

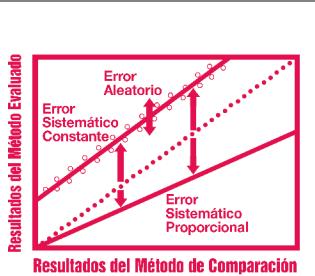
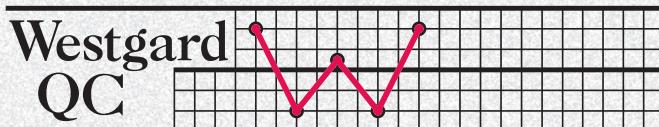
Validación Básica de Método



Edición
Wallace
Coulter

Entrenamiento en Gestión
de la Calidad Analítica
para Laboratorios Clínicos

James O. Westgard, Ph.D.



Tipo de Error Analítico	Experimentos de Evaluación	
	Preliminares	Finales
Error Aleatorio	Replicación Intra corrida	Replicación Entre corridas
Error Sistématico Constante	Interferencia	Comparación de Métodos
Error Sistématico Proporcional	Recuperación	



Validación Básica de Método

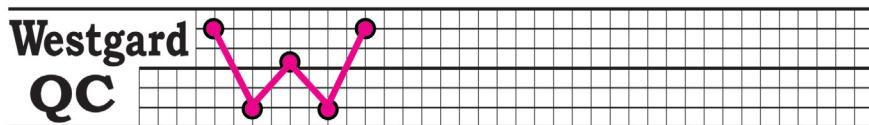
EDICIÓN WALLACE COULTER

James O. Westgard, PhD

con la colaboración de
Patricia L. Barry, BS, MT (ASCP)
R. Neill Carey, PhD, FACB
Sharon S. Ehrmeyer, PhD
Elsa F. Quam, BS, MT (ASCP)
Sten Westgard, MS

Gabriel A. Migliarino Bioq.

Copyright © 1999, 2003, 2008, 2013
7614 Gray Fox Trail, Madison WI 53717
Teléfono 608-833-4718 HTTP://WWW.WESTGARD.COM



Biblioteca del Congreso Número de control: 2013913408

**ISBN 978-1-886958-18-0
Publicado por QC Westgard, Inc.
7614 Gray Fox Trail
Madison, WI 53717**

teléfono 608-833-4718

Copyright © 1999, 2003, 2008, 2013 por QC Westgard, Inc. (WQC). Todos los derechos reservados. Ninguna parte de esta publicación puede ser reproducida, almacenada en un sistema de recuperación, o transmitida en cualquier forma o por cualquier medio, ya sea electrónico, mecánico, fotocopia, grabación, o de otra manera, sin el permiso previo por escrito de QC Westgard, Inc. Esta edición está financiada por una beca de la Fundación Wallace Coulter.

Dedicatoria



Wallace Henry Coulter February 17, 1913 - August 7, 1998

Introducción

Wallace H. Coulter fue un prolífico inventor, un innovador y emprendedor. Aunque algunas personas pueden no saber su nombre, sus invenciones e ideas han afectado prácticamente a todo el mundo.

El Principio de Coulter (USPTO # 2656508), desarrollado en 1948, es una de sus más de 80 patentes, muchas de las cuales fueron invenciones realizadas años más tarde. El principio de Coulter o "zona de detección electrónica" es un método para el recuento y medición de las partículas suspendidas en un fluido. El Principio Coulter trajo grandes avances en ciencia, medicina y en la industria. De hecho, todavía toca la vida de todos hoy, al realizar un análisis de sangre, para pintar su casa, beber cerveza, comer chocolate, tragarse una pastilla o aplicarse cosméticos. Es fundamental para toners y cerámicas, así como también en la exploración del espacio.

Nacido en Little Rock y criado en McGehee, Arkansas, Wallace era un hombre humilde integro, impulsado por la filosofía de la ciencia al servicio de la humanidad. Con su hermano Joseph R. Coulter, Jr., Wallace fundó la Corporación Coulter y lideró esta empresa privada global de diagnóstico durante toda su historia de 40 años. En 1997, *Beckman Instruments* adquirió la Corporación Coulter. Cuando Wallace falleció en agosto de 1998, su patrimonio fundó la Fundación Wallace H. Coulter. La Fundación continúa su misión trabajando con dieciséis universidades líderes de Estados Unidos implementando procesos de desarrollo de productos usados en la industria médica cuyo propósito es el de acelerar y aumentar el número de invenciones e innovaciones desarrolladas en las universidades que se convierten en productos que benefician pacientes. Cien años después de su nacimiento, y más de una década después de su muerte, este héroe americano deja atrás el legado de sus logros, así como colegas, socios, amigos y familiares en todo el mundo que se inspiraron en su influencia y continúan con su misión de la ciencia al servicio de la humanidad.

"Decir poco y lograr mucho demuestra las características de un gran hombre."

Wallace Coulter

The Owl-1929 High School Yearbook, McGehee, AR

El Hombre

Wallace Coulter creció durante la era de la invención. Nacido en febrero de 1913, Wallace pasó sus primeros años en McGehee, Arkansas, un pequeño pueblo cerca de Little Rock. Wallace siempre tuvo una mente inquisitiva. A los tres años, quedó fascinado con los números y los dispositivos. Cuando se le ofreció una bicicleta para su undécimo cumpleaños, él solicitó en su lugar su primer dispositivo de radio. Después de graduarse de la escuela secundaria a los 16 años, asistió a su primer año de universidad en el Westminster College en Fulton, Missouri, sin embargo, su interés en la electrónica lo impulsó a transferirse al Instituto de Tecnología de Georgia para su segundo y tercer año de estudio. Esto fue a principios de 1930, y debido a la Gran Depresión, no pudo completar su educación formal.

El interés de Wallace en la electrónica se manifestó en una variedad de puestos de trabajo no convencionales. Por ejemplo, trabajó para la radio WNDR en Memphis, TN como locutor auxiliar, en man-

tenimiento de equipos y en la realización de algunos de los primeros experimentos en comunicaciones móviles. En 1935, se unió a General Electric X-Ray como un ingeniero de ventas y servicio en el área mantenimiento de equipos médicos en Chicago. Aunque la práctica de las empresas estadounidenses de emplear expatriados no era común antes de la Segunda Guerra Mundial, cuando la posibilidad de cubrir el Lejano Oriente estuvo disponible, Wallace aprovechó la oportunidad de vivir y trabajar en el extranjero. Durante los siguientes veinticuatro meses Wallace se basó en Manila, Shanghai y Singapur, brindando servicios a todo el Extremo Oriente. Admiraba el exuberante paisaje y la variedad de frutas tropicales y quedó fascinado con la historia China, su arte y su cultura. Wallace estaba en Singapur cuando los Japoneses amenazaron la ciudad a finales de 1941. En diciembre, cuando los Japoneses comenzaron a bombardear la ciudad, se encontró con un pequeño barco de carga con destino a la India y se fue al amparo de la oscuridad. Después de unas semanas en la India, Wallace se dio cuenta de que el regreso a los Estados Unidos a través de Europa era imposible. Eligió un camino a casa no tan directo, iniciando su travesía a través de África y América del Sur. Le tomó cerca de 12 meses, por fin regresar a los EUA en la Navidad de 1942. La estancia de Wallace en el Lejano Oriente y su larga travesía a casa atravesando cuatro continentes, fue una experiencia transformadora para un joven de una pequeña ciudad de Estados Unidos de América. Influyó para siempre sus valores, tanto profesionales como personales.

Aunque Wallace Coulter recibió muchos reconocimientos, siempre evitó la publicidad y ser el centro de atención. Para Wallace, el logro fue importante y no los elogios. Vivía modestamente e invirtió una parte extraordinaria de las ganancias de la compañía de nuevo en la investigación y el desarrollo. Wallace permaneció solo durante toda su vida, su empresa y sus empleados se convirtieron en su familia. Wallace era un hombre compasivo que siempre animó a sus empleados a soñar y hacer lo mejor. Él ayudó personalmente a muchos de los empleados y a sus familias, proporcionando préstamos, becas y patrocinios. Tras la venta de la Corporación Coulter, Wallace se aseguró de que su familia de empleados se beneficiara mediante la asignación de un fondo total de \$ 100 millones para que se pagara, en función de sus años de servicio, a cada uno de los empleados en todo el mundo.

"Cuando Joe y yo estábamos perfeccionando el Principio de Coulter, difícilmente pudimos haber soñado con la explosión tecnológica que se derivaría de ello."

Wallace Coulter
1995

El Inventor

Después de la guerra, Wallace trabajó para varias empresas de electrónica, incluyendo Raytheon y Mittleman Electronics en Chicago. Como todo gran inventor, Wallace mantiene un laboratorio de garaje en su casa para trabajar con ideas y proyectos prometedores.

Trabajando en un proyecto para el Departamento de Investigación Naval, Wallace estaba tratando de estandarizar el tamaño de partículas de pigmento utilizados en la pintura en los acorazados de Estados Unidos con el fin de mejorar su adherencia al casco. Una fría noche de invierno tempestuoso, Wallace encontró que el suministro de pintura para el experimento se había congelado mientras él estaba fuera. Como no quería volver a la intemperie, se preguntó: "¿Qué sustancia tiene una viscosidad similar a la pintura y esta fácilmente disponible?" Usando su propia sangre, una aguja y un poco de celofán, inventó lo que más tarde se conoció como el Principio Coulter® – el principio de la utilización de impedancia eléctrica para contar y medir partículas microscópicas suspendidas en un fluido.

Los primeros intentos de Wallace de patentar su invento fueron rechazados por más de un abogado que creía que "no se puede patentar un agujero". Persistente como siempre, Wallace finalmente solicitó su primera patente en 1949 y se emitió (USPTO # 2656508) el 20 de octubre de 1953. El Principio Coulter es responsable de gran parte de las prácticas actuales en los laboratorios clínicos en el área de hematología. Hoy en día y todavía después de 60 años, la mayoría de los recuentos sanguíneos completos (CBC, por sus siglas en inglés) se realizan en instrumentos utilizando el Principio Coulter.

"Todos los problemas tienen solución... sólo debemos romperlos en trozos pequeños y resolver de una sola pieza a la vez"
Wallace Coulter

El innovador

Las dos grandes pasiones de la vida de Wallace eran aplicar los principios de la ingeniería para ayudar a resolver los desafíos clínicos y abrazar así a la diversidad de las culturas del mundo. Su primera pasión lo llevo a inventar el Principio Coulter™. Su segunda pasión lo llevó a establecer más de veinte subsidiarias internacionales. Luego esto se transformó en una práctica estándar de negocios, Wallace identificó que era imperativo emplear personal local para servir a sus clientes.

Cuando era llamado a desempeñar tareas en hospitales trabajando para GE, Wallace observo a los trabajadores del laboratorio inclinados sobre los microscopios contando manualmente las células sanguíneas desparramadas sobre porta objetos. Práctico como siempre, y sabiendo que debía existir una forma más fácil de realizar esa tarea, Wallace focalizó la primera aplicación del Principio Coulter en el recuento de glóbulos rojos. Este instrumento fue conocido como el Contador™ Coulter®. Mediante la automatización de este proceso laborioso que se realizaba con anterioridad de forma manual, Wallace revolucionó la práctica de la hematología en el laboratorio clínico. A través de lo que se ha convertido en la prueba de diagnóstico más solicitada en el mundo entero, el recuento completo de células sanguíneas (CBC, por sus siglas en inglés), Wallace ayudó a cambiar de manera radical la relación médico paciente. Wallace también fue pionero en el campo de la citometría de flujo y anticuerpos monoclonales. Estas tecnologías son utilizadas en la caracterización y tratamiento del cáncer, leucemia y enfermedades infecciosas. El anticuerpo B-1 (anti-CD20, por sus siglas en inglés), comercializado como Bexxar, desarrollado bajo su dirección está aprobado por la FDA para pacientes con linfoma folicular para pacientes con recaída o las cuales la enfermedad persiste después del tratamiento inicial.

"Los desafíos son buenos, y verdaderamente teníamos una gran cantidad de ellos"
Wallace Coulter

El empresario

En 1958, Wallace y su hermano, Joseph R. Coulter, Jr., fundaron Coulter Electronics para fabricar, comercializar y distribuir sus Contadores Coulter. Desde los comienzos fue una empresa familiar, con el Sr. Joseph, su padre, desempeñándose como secretario tesorero. Wallace y Joe, Jr. construyeron los primeros modelos, los cargaron en sus carros y vendieron personalmente cada unidad. En 1959, para proteger los derechos de patente, se establecieron subsidiarias en el Reino Unido y Francia. En 1961, los hermanos Coulter trasladaron su empresa en crecimiento al área de Miami, Florida, donde ellos permanecieron por el resto de sus vidas.

Wallace insistió en que la compañía debía mantener la cadena completa de suministro: investigación y desarrollo, fabricación, distribución, ventas financiación, entrenamiento y servicio y soporte post venta. De hecho la Corporación Coulter fue la primera compañía de diagnóstico que daba soporte a sus instrumentos con los reactivos y materiales de control necesarios para la operación completa del sistema.

Bajo el mando de Wallace como presidente, la compañía se convirtió en la industria líder en equipamiento para el análisis de células sanguíneas, empleando casi 6.000 personas en 19 países, con más de 50.000 instalaciones de instrumentos. La compañía ha dado lugar a una familia entera de instrumentos, reactivos y controles, no solamente en el área de hematología sino también para citometría de flujo, análisis industrial de partículas finas y otros diagnósticos de laboratorio.

En la década de los 90, la Corporación Coulter era una de las más grandes empresas privadas de diagnóstico en el mundo. En Octubre de 1997, la Corporación Coulter fue adquirida por *Beckman Instruments, Inc.*, y la compañía pasó a ser conocida como *Beckman Coulter, Inc.*

“Vamos a hacerlo más fácil, más rápido y de manera más confiable que nunca para diagnosticar enfermedades. El paciente será el gran beneficiado.”

Wallace Coulter

El legado

Wallace fue una persona muy reservada que no buscó reconocimiento público, pero sus logros son numerosos. Recibió alrededor de 80 patentes, muchas de las cuales le fueron asignadas por los descubrimientos realizados años después de la invención del principio Coulter®. Wallace Coulter fue galardonado con el prestigioso premio John Scott por su labor científica que se les otorga a inventores cuyas innovaciones han tenido un efecto revolucionario sobre la humanidad. Con este premio creado en 1816 para “los hombres y las mujeres ingeniosas”, Wallace se unió a sus héroes de la infancia; Thomas Edison, Maria Curie, Jonas Salk y Guillermo Marconi.

Recibió doctorados honorarios del Georgia Institute of Technology, *Westminster College*, la Universidad de Clarkson, la Universidad de Miami y la Universidad Barry. Fue honrado con el Premio IEEE Morris E. Leeds, Industrial del Año Florida, y el *Buylne Samme Lifetime Achievement Award MD*. Wallace es la única persona que no siendo médico o un hematólogo ha recibido por parte de la Sociedad Americana de Hematología (ASH, por sus siglas en inglés) el Premio al Servicio Distinguido por su enorme contribución al campo de la hematología. Él fue el destinatario del *Association of Clinical Scientist's Gold Headed Cane Award* y miembro del Instituto Americano de Ingeniería Médica y Biológica (AIMBE, por sus siglas en inglés). En 1998, fue incluido en la Academia Nacional de Ingeniería. En 2004, Wallace fue incluido en el *National Inventor's Hall of Fame*. En el 2013, año de centésimo aniversario fue reconocido como el primer recipiente del “Golden Goose Award” que reconoce a científicos e ingenieros los cuales a través de investigaciones fundadas por el gobierno Americano han tenido un profundo impacto para la humanidad y creado gran beneficio económico para la nación.

Cuando Wallace falleció en agosto de 1998, donó toda su riqueza para la formación de la Fundación Wallace H. Coulter. La Fundación continúa su legado, trabajando con las principales universidades de Estados Unidos para establecer un proceso de tipo empresarial para acelerar la introducción de las innovaciones académicas del área de la salud en el mercado con el objetivo de mejorar la atención al paciente. Además, en reconocimiento a la pasión de Wallace para mejorar la atención sanitaria en los países en desarrollo, la Fundación está trabajando en colaboración con las sociedades médicas de Estados Unidos para llevar a los últimos avances en las prácticas médicas y de laboratorio para estos países. De especial interés es el trabajo de la Fundación en América Latina con la Asociación Americana de Química Clínica (AACC, por sus siglas en inglés) y la Sociedad Americana de Hematología (ASH, por sus siglas en inglés). En el tercer año de un compromiso de ocho años, con el objetivo de llegar a 5.000 profesionales de América Latina, el objetivo de la iniciativa juntamente con la AACC es elevar la práctica de la calidad analítica en los laboratorios clínicos. Trabajando con la Sociedad Americana de Hematología (ASH siglas en Ingles) la Fundación, en 2007, se estrenó los *Highlights of ASH®* para ofrecer programas educativos sobre los avances en los tratamientos clínicos para los pacientes con cáncer de la sangre y otros trastornos sanguíneos. El programa tuvo tanto éxito que se ha convertido en un encuentro anual.

Ahora, cien años después de su nacimiento, y más de una década después de su muerte, este héroe americano deja atrás el legado de sus logros, así como colegas, socios, amigos y familiares en todo el mundo que se inspiraron en su influencia y continúan su misión de lograr que la ciencia esté al servicio de la humanidad.

Prólogo a la edición Wallace Coulter

James O. Westgard, PhD

Validación Básica de Método es parte de una trilogía de libros “de vuelta a los principios básicos” que se enfocan en el manejo de la calidad analítica. Los otros dos libros son Prácticas Básicas de Control de la Calidad y Planificación Básica para la Calidad. Cuando yo enseño estos materiales, comienzo con la validación de métodos ya que introduce los conceptos básicos del desempeño analítico y las técnicas estadísticas y experimentales que se necesitan para describir el desempeño en términos cuantitativos. Estos conceptos conllevan la práctica de Control Estadístico Interno de la Calidad y la selección de procedimientos de Control de la Calidad óptimos a través del diseño y la planificación de la calidad.

La fuente original de este enfoque de la validación de métodos retrocede treinta años en el tiempo a una serie de documentos que fueron publicados en el “*American Journal of Medical Technology*” y más tarde como una monografía titulada “*Method Evaluation*”. Mis co-autores eran Diane J de Vos, Marian R. Hunt, Else F. Quam, R. Neill Carey, y Carl C. Garber, todos los cuales trabajaban en la Universidad de Wisconsin. Introdujimos este enfoque en los talleres que dábamos en los encuentros nacionales ASMT y AACC. Hoy en día, Niell y Carl, junto con David Koch, continúan enseñando este enfoque en el encuentro nacional AACC. Ahora ellos tienen el record del taller más largo en la historia de AACC.

Esta edición brinda importantes actualizaciones basadas en nuevos requisitos regulatorios y estándares de buenas prácticas, particularmente los últimos lineamientos de CLSI (*Clinical and Laboratory Standards Institute*):

- Regulaciones y requisitos de acreditación actualizados de CLIA;
- Un capítulo revisado sobre el rango reportable que incluye la verificación de la calibración;
- Un capítulo revisado sobre límites de detección que incluye los conceptos de Límite en Blanco, Límite de Detección y Límite de Cuantificación, como se recomienda en los lineamientos de la guía CLSI EP17 A;
- Una Carta de Decisión de Método actualizado que incluye criterios para el desempeño 2, 3, 4, 5 y 6-Sigma;
- Un nuevo capítulo acerca de la estimación de la veracidad y precisión basada en los lineamientos de la guía CLSI EP15 – A2, que incluye instrucciones para realizar los cálculos requeridos utilizando hojas de cálculo electrónicas;
- Un nuevo capítulo acerca de los estándares globales en desarrollo (ISO, *International Standards Organization*), ISO 15189 y los conceptos de veracidad e incertidumbre de medición;
- Un nuevo capítulo acerca de las métricas Seis Sigma, que incluyen instrucciones para convertir los datos de la validación de método a la métrica Sigma.

Por más de treinta años, he trabajado en el control de la calidad y la validación de métodos. Mientras que las estadísticas, las ecuaciones y los cálculos no cambian, el contexto y el entorno están en constante evolución. Espero que esta tercera edición los ayude a entender estos números de validación de métodos en el contexto adecuado de su laboratorio.

James O. Westgard
Madison Wisconsin

Agradecimientos

En primer lugar, Sten Westgard merece el crédito por la “fórmula” utilizada para desarrollar estos materiales. Él es el responsable de la publicación de estos materiales y también brinda una fuerza impulsora para su finalización. A mi edad, necesito fechas límite para finalizar las cosas y las conseguí gracias a Sten.

Varios colegas me han ayudado en el desarrollo de estos materiales, incluyendo a Elsa Quam y Patricia Barry quienes han sido mis socias por largo tiempo en los Laboratorios Clínicos de la Universidad de Wisconsin Hospital y Clínica. Sharon Ehrmeyer continúa ayudando a clarificar el maravilloso mundo de las reglas CLIA, sus regulaciones y lineamientos de acreditación e inspección relacionados. Neill Carey contribuyó con su experiencia y conocimiento para hacer que la presentación de los lineamientos del EP15 resultara práctica y útil.

Los mapas antiguos que aparecen en este libro son parte de una pequeña colección personal. Espero que los encuentren útiles para ilustrar las ideas principales en el libro, así como también interesantes y bellos documentos históricos.

En la Fundación Wallace Coulter, queremos agradecer por sus incansables esfuerzos y apoyo a Elias Caro , Mara Neal y Sue Van. Sin ellos este libro no hubiese existido.

Autores y colaboradores

James O. Westgard, PhD, Profesor Emérito en el Departamento de Patología y Medicina de Laboratorio de la Universidad de Wisconsin, escuela de Medicina, donde continúa enseñando en el programa de Ciencias de Laboratorio Clínico. Él también es el presidente de Westgard Quality Control Inc. Su biografía completa puede ser encontrada en <http://www.westgard.com/jimbio.htm>

Patricia L. Barry, BS;MT (ASCP), Especialista en calidad en los laboratorios clínicos de la Universidad de Wisconsin en hospitales y clínicas.

R. Neill Carey, PhD, FACB, Director de Química Especial en el Centro Médico Regional Península, Salisbury, Maryland.

Elsa F. Quam, BS, MT (ASCP), Especialista de la calidad en los laboratorios clínicos de la Universidad de Wisconsin en hospitales y clínicas.

Sten Westgard, BA, MS, Director de Servicio al Cliente y Tecnología de Westgard QC.

Gabriel A. Migliarino Bioq., Director de Gmigliarino Consultores, Especialista de la Calidad en laboratorios Clínicos en América Latina, consultor para la implementación de la Norma ISO 15189, trabajando actualmente en varios países del área.

Hay más en línea en “Westgard Web”

Para hacer lugar a todas las actualizaciones, revisiones y capítulos nuevos por completo dentro del libro, sin que fuera demasiado extenso, tuvimos que hacer algunos recortes. Notablemente, tuvimos que recortar el glosario y las listas de referencia.

Pero no se preocupe, puede ver estas fuentes si las necesita.

Puede ir a <http://www.westgard.com/bmv/extras.html> para el acceso y los enlaces a estas funciones:

- Preguntas frecuentes (FAQs, por sus siglas en inglés)
- Glosario de términos
- Lista completa de referencia para este libro
- Enlaces a hojas de cálculo, hojas de trabajo y otras descargas
- Enlaces a calculadoras para Validación de Métodos, incluyendo algunas calculadoras que no están disponibles para el público en general.

Tabla de Contenidos

CAPITULO 1: ¿Es la calidad aún un asunto relevante para los exámenes de laboratorio?	1
CAPITULO 2: ¿Cómo gestiona usted a la calidad?	13
CAPITULO 3: ¿Cuál es el propósito de un estudio de validación de métodos?	27
CAPITULO 4: ¿Cuáles son los requisitos regulatorios para la validación de métodos?	37
CAPITULO 5: ¿Cómo es seleccionado un método?	53
CAPITULO 6: ¿Qué experimentos son necesarios para validar el desempeño de un método?	63
CAPITULO 7: ¿Cómo se analizan los datos experimentales?	75
CAPITULO 8: ¿Cómo se calculan estas estadísticas?	87
CAPITULO 9: ¿Cómo se determina el rango reportable de un método?	103
CAPITULO 10: ¿Cómo se determina la imprecisión de un método?	115
CAPITULO 11: ¿Cómo se determina la inexactitud o sesgo de un método?	127
CAPITULO 12: ¿Cómo usar la estadística para estimar errores analíticos?	139
CAPITULO 13: ¿Cómo examinar fuentes específicas de inexactitud?	157

CAPITULO 14: ¿Cuál es el valor más bajo confiable para un método?	173
CAPITULO 15 ¿Cómo se verifica el intervalo de referencia de un método?	185
CAPITULO 16: ¿Cómo evalúa el desempeño de un método?	195
CAPITULO 17: ¿Cuál es un procedimiento práctico para validar un método?	207
CAPITULO 18: ¿Cómo usa la estadística en el Mundo Real?	221
CAPITULO 19: ¿Cómo pueden ser verificadas las especificaciones de desempeño de los fabricantes?	231
CAPITULO 20: ¿Cómo pueden ser evaluadas las especificaciones de desempeño declaradas por los fabricantes en una escala Sigma?	251
CAPITULO 21: ¿Qué impacto tendrá ISO en la gestión de la calidad analítica?	263
CAPITULO 22: Respuestas de Auto Evaluación	285
APÉNDICE 1: CALIDAD ANALÍTICA CLIA '88 REQUISITOS	313
ÍNDICE	317

1: ¿Es la calidad aún un asunto relevante para los exámenes de laboratorio?

En este capítulo introductorio, el Dr. Westgard desafía las ideas actuales de que la calidad analítica es mejor que la necesaria para el cuidado de la salud. Haciendo uso de mapas históricos reconocidos como los registros más fiables y exactos de esos tiempos, ilustra que las “verdades” populares pueden estar bien documentadas y tener amplio crédito, y aún así ser erróneas. Él afirma la necesidad de definir los requisitos de la calidad para manejar la calidad en forma cuantitativa.

Objetivos:

- Comenzar a pensar en la calidad de forma crítica.
- Reconocer que las creencias actuales sobre la calidad podrían no estar efectivamente fundamentadas.
- Identificar los puntos críticos en el manejo de la calidad en forma cuantitativa.

Materiales del capítulo:

- **Mitos de la calidad**, por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar la lección.
- Encontrar qué calidad es necesaria para una prueba de colesterol.

Preguntas de Autoevaluación:

- ¿Por qué existen hoy en día mitos sobre la calidad en el laboratorio clínico?
- ¿Qué se puede hacer para mejorar las prácticas de gerenciamiento de la calidad en el laboratorio clínico?



¡La isla mítica de California!

NOUEAU MEXIQUE ET CALIFORNIE, por Alan Mallet, Paris 1686. La miniatura de un mapa francés que muestra California como una isla plana... un mito que persistió desde 1620 por más de 100 años.

Mitos de la Calidad

James O. Westgard, PhD

¡UN MITO es una Historia, Teoría o Hipótesis Equivocada!

Mitos Históricos de la Cartografía

Isla mítica de California. ¿Sabías que California fue una isla? Está bien documentado en los mapas de mejor reputación del siglo XVII que California estaba completamente rodeada de agua. Por ejemplo, vea el mapa adjunto que muestra la *Isle de Californie*. Ahí está, documentado en blanco y negro, prueba de que California fue una isla.

Este mapa de *Nouveau Mexique et Californie* hecho por Alain Mallet fue publicado en 1686 en la *Description de l'Univers* (París). Mallet copió el modelo de California que apareciera en un mapa anterior hecho por Sanson, que era uno de los más distinguidos cartógrafos franceses (era muy común que los cartógrafos se copiaran los trabajos unos a otros). Cuando aparecía un nuevo descubrimiento en un mapa, era ampliamente “diseminado” en los otros mapas de la época. El descubrimiento de que California era una isla fue documentado por primera vez en 1622 y persistió en cartas geográficas hasta 1750, a pesar de que la evidencia en 1705 claramente establecía que esto era un mito.

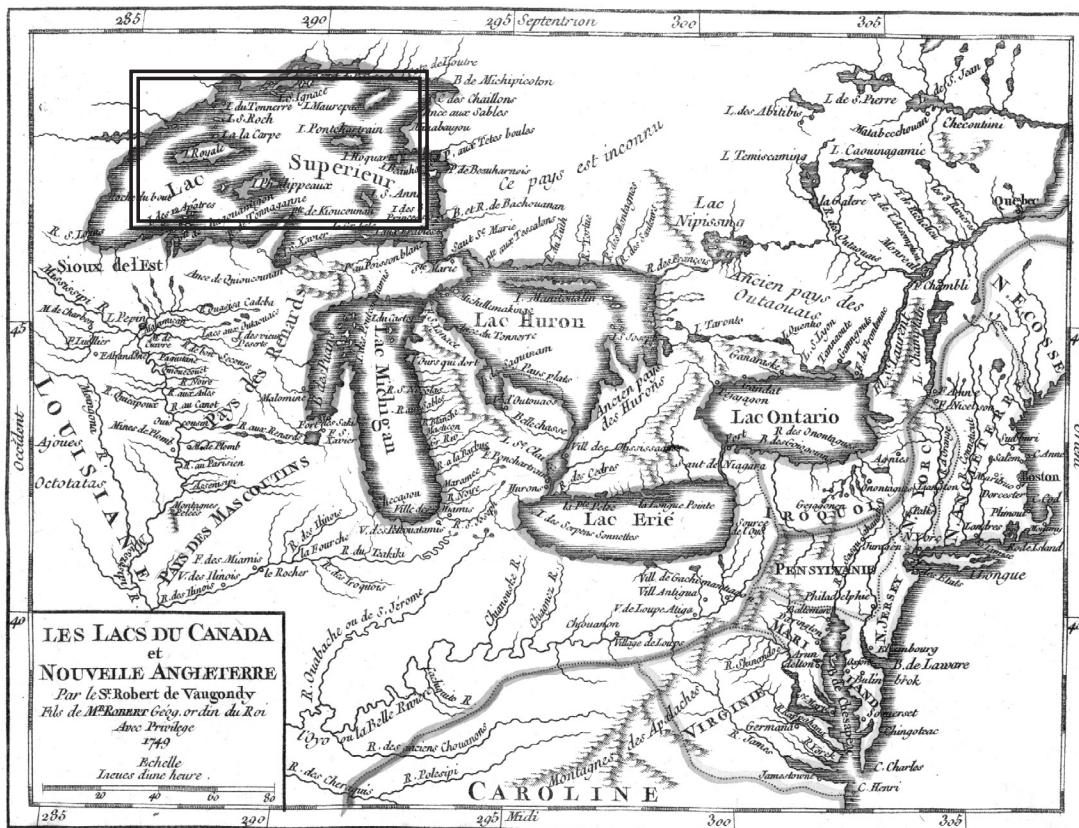
Isla mítica de Friesland. De hecho, hay toda una historia sobre las islas míticas, lo que sugiere que estos mitos no son tan atípicos como uno podría esperar. A finales del siglo XVI, uno de los más famosos cartógrafos, Abraham Ortelius, realizó un mapa del Atlántico Norte que mostraba la isla de Friesland que yacía a unos pocos kilómetros al suroeste de Islandia, complementado con una descripción detallada de la zona costera, los puertos, la gente que vivía ahí, su aspecto físico y su *modus vivendi*. Es un mapa hermoso, decorado con barcos zarpendo y especies marinas, y era el mapa más reconocido del área en ese tiempo. El único problema era que Friesland no existía. Cuando la gente zarpaba hacia el nuevo mundo y pasaba Islandia, adscribían más y más detalles y daban mayor realidad a Friesland porque suponían que era el siguiente cuerpo de tierra.



¡La isla mítica de Freisland!

SEPTENTRIONALES REGIONES, por Philip Galle, Amberes, 1595. Una miniatura de un famoso mapa de Ortelius de la reion del Atlantico norte que muestra la mítica isla de Friesland (ver cuadro) localizada al suroeste de Islandia.

Isla mítica en el Lago Superior. Otro ejemplo que es de interés a aquellos de nosotros que somos del medio oeste norteamericano son las islas de Ponchartrain y Phillippeaux en el Lago Superior. Cuando la frontera entre los Estados Unidos y Canadá fue establecida por el Tratado de París en 1783 se decidió que estas islas fueran parte de E.U.A. En los inicios del siglo XIX, cuando se estaba estableciendo Wisconsin, el gobierno de E.U.A. envió exploradores a cartografiar el área de forma más completa, ¡pero no pudieron encontrar estas islas! Se mostraban en todos los mapas de la época, pero no aparecían sobre el agua. Al parecer, los exploradores crearon estas islas y les dieron el nombre del Ministro de Gobernación que estaba patrocinando sus investigaciones. Probablemente estaban tratando de conseguir más dinero para exploraciones futuras y necesitaban algunos descubrimientos preliminares que justificaran esto.



¡La isla mítica en el Lago Superior!

LES LACS DU CANADA et NOUVELLE ANGLETERRE, por Robert de Vaugondy, Paris, 1749.

Éste mapa muestra el Lac Supérieur con la verdadera Isla Royal y las islas míticas de Phillipaux, Pontchartrain, Maurepas, St. Anne.

Mitos modernos de la calidad

Estos mitos cartográficos son divertidos en retrospectiva, pero fueron tomados muy seriamente en la época y crearon problemas posteriores. Existen mitos hoy en día que también se toman seriamente y que causarán serios problemas en el futuro. Algunos de ellos repercuten muy cerca de casa... la calidad del cuidado de la salud y la calidad de los exámenes de laboratorio.

Mito: El Aseguramiento de la Calidad (AC) asegura la calidad en el cuidado de la salud. Es un cuento erróneo que pone un buen giro a los esfuerzos actuales de medir la calidad en el cuidado de la salud. Como profesionales de la salud, hablamos de aseguramiento de la calidad (AC), pero nuestros programas de aseguramiento de la calidad (que usualmente son requisitos para la regulación y la acreditación) se enfrentan en primera instancia con el desempeño de las mediciones. La Evaluación de la Calidad sería un mejor nombre para estos esfuerzos. En tanto que es importante evaluar la calidad para saber qué tan bien vamos, medir la calidad no **asegura** que la calidad necesaria será

alcanzada. Lograr la calidad requiere, de hecho, de un diseño y una planificación de la calidad, que comienza con la definición de la calidad que se necesita, para luego construir esa calidad en el proceso.

Mito: El Control Estadístico Interno de la Calidad (CC) controla la calidad de los exámenes de laboratorio. Es una *teoría errónea* que el uso de la estadística de algún modo asegura que los resultados de los exámenes del laboratorio tienen la calidad necesaria. Virtualmente todos los laboratorios aplican Control Estadístico Interno de la Calidad (CC) como parte de sus esfuerzos para asegurar la calidad de los exámenes de laboratorio. Mientras que podríamos no entender la teoría o la estadística, aún creemos que algo mágico ocurre gracias a esa estadística. De algún modo, al analizar los controles y graficar los resultados en cartas de control, esperamos controlar la calidad de nuestros procesos de examen, incluso aún si no entendemos cómo funciona.

Mito: La Calidad puede ser controlada incluso si no se conocen la calidad requerida. Es una *hipótesis errónea* que la calidad puede ser controlada cuando no sabemos cuál es la calidad necesaria. Muy pocos laboratorios han definido la calidad analítica necesaria para los exámenes que realizan. ¿Cómo es posible que sepamos que estamos logrando lo desconocido? ¿Puede usted manejar las finanzas del laboratorio sin conocer el presupuesto? ¿No necesita saber los requisitos de la calidad para un examen de laboratorio si piensa controlar la calidad de los procesos de examen?

Mito: Los requisitos de la calidad necesitan considerar sólo a la imprecisión y a la inexactitud. Este problema con los requisitos de la calidad resulta incluso más complicado. Las ideas actuales acerca de las metas y los requisitos de la calidad son erróneas porque consideran sólo las características de desempeño del método estable (imprecisión e inexactitud). Si el desempeño es estable, ¿por qué molestarnos en hacer control de la calidad en absoluto? Si el CC es necesario, ¿no tenemos que considerar las características de desempeño de los procedimientos de CC (probabilidad de detección de errores y de falsos rechazos) en los esquemas a lograr?

Mito: Los métodos de laboratorio actuales tienen mejor imprecisión e inexactitud que la necesaria. El efecto neto de todos estos mitos es la creencia de que el desempeño de los métodos actuales de laboratorio es mejor que el que exigen las necesidades médicas. Esta creencia está basada en una teoría errónea de establecimiento de metas de la calidad, una hipótesis errónea al igualar todas las variaciones médicas tolerables con la variación analítica, sin tomar en cuenta la variabilidad biológica del propio sujeto, y la presunción de que los procedimientos de CC tienen curvas de respuesta ideales que pueden detectar cualquier cambio en el desempeño, sin importar cuán pequeño sea.

Mito: La calidad analítica se da por sentado hoy en día. Como consecuencia de estos mitos, hay un sentimiento común hoy en día de dar por sentado a la calidad analítica, esto es, que la calidad analítica por sí misma es asumida hoy en día. En el medio de programas Seis Sigma, Lean, Manejo de Riesgos, y Manejo Total de la Calidad (MTC), es a menudo asumido erróneamente que los problemas en el manejo técnico de la calidad ya han sido resueltos. Esto representa la hipótesis errónea de que los esfuerzos en el pasado han resuelto todas las dificultades técnicas, así que ahora podemos pasar a nuevos asuntos que estén en boga, tales como el monitoreo de la satisfacción del usuario, la medición de los resultados de pacientes, etc.

Mito: No se necesitan otras mejoras en la calidad analítica. El resultado colectivo de todos estos mitos es un falso sentimiento de seguridad respecto a la calidad de los procesos de examen del laboratorio. Muchos piensan que la calidad analítica es tan buena que no se necesitan nuevas mejoras. Este es el mito más grave de todos porque nos hace condescendientes con lo que estamos haciendo y obstaculiza los esfuerzos de mejorar a la calidad analítica de los exámenes de laboratorio.

Mito: El gobierno regula los exámenes de laboratorio para asegurarse de que la calidad es aceptable. Si bien es cierto que la FDA (*Food and Drug Administration*) aprueba los nuevos exámenes y sistemas analíticos, es importante comprender que esta depuración está basada solamente en la “verdad en el etiquetado”. Se solicita a los fabricantes presentar pruebas sobre ciertas características de desempeño, como el rango reportable, precisión, exactitud, interferencia, límite de detección y intervalos de referencia, y referir los datos que respalden esas pruebas. Los procesos de la FDA se enfocan en si los datos confirman o no las pruebas presentadas por los fabricantes, no en si la calidad del proceso de examen es o no aceptable para el cuidado del paciente. Podríamos creer que los fabricantes no presentarían un nuevo examen para la depuración de la FDA a menos que la calidad fuera aceptable, pero esa presunción no siempre es correcta.

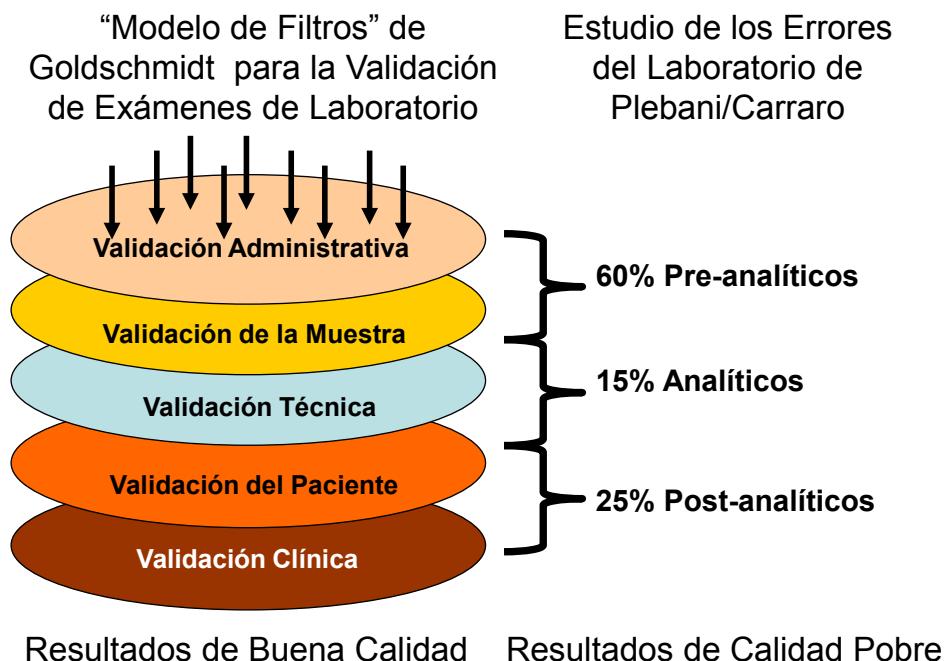
Mito: Los laboratorios deberían enfocarse hoy en día en los errores pre-analíticos y post-analíticos debido a que los errores analíticos ya no son un problema. Esta idea afloró en 1990 del CDC (*Centers for Disease Control and Prevention*) en un esfuerzo para ampliar los esfuerzos de la evaluación de la calidad en los laboratorios clínicos. Más tarde en los 90, CMS (*Centers for Medicare & Medicaid Services*) adoptó esa perspectiva en la revisión de las Reglas Finales de CLIA (*Clinical Laboratory Improvement Amendments*) para incluir el manejo de la calidad del “proceso total de examen”, esto es, las partes pre-analítica, analítica y post-analítica de la metodología de examen. Para los inicios del siglo XXI, esta creencia fue ampliamente adoptada en la comunidad de los laboratorios clínicos y hoy en día es usada (y aceptada) como un argumento

para reducir la cantidad de Control Estadístico Interno de la Calidad aplicada durante la fase analítica. Satisface nuestro anhelo de hacer menos Control Estadístico Interno de la Calidad, simplificar los exámenes de laboratorio y reducir los costos y eliminar las repeticiones y de resolución de problemas, todo lo cual permite a los exámenes de laboratorio ser realizados en sitios de examen donde las habilidades técnicas y la experiencia en el laboratorio pueden faltar.

¿Dónde está la evidencia?

En esta época de Medicina Basada en Evidencia de Laboratorio, ¿dónde están los datos que apoyan estas creencias de que no necesitamos ya preocuparnos por la calidad analítica? Mientras que tomaría muchas páginas de este libro desaprobar este conjunto entero de creencias equivocadas, consideremos éste último mito (los laboratorios se deben enfocar en las fases pre-analítica y post-analítica del proceso total de examen) porque descansa en muchas otras creencias. De hecho, sería el enfoque correcto si los errores analíticos no fueran más un problema.

Primero, consideremos un modelo más completo del proceso total de examen, como provee Goldschmidt y cols. (1), mostrado en la figura abajo. Llamado “modelo de filtros”, la figura ilustra una serie de filtros a través de los cuales una petición de un examen de laboratorio, un espécimen y una muestra deben pasar. En la realidad, estos son filtros matemáticos más que filtros de papel como se sugiere en la figura, pero como científicos de laboratorio más que matemáticos, el enfoque es fácil de comprender en la ilustración. Este modelo más detallado describe 5 fases o filtros para la validación del proceso total de examen.



- La validación administrativa se refiere a los pasos que comienzan con la selección y solicitud del examen correcto, la recolección de la información correcta para comprender el contexto del examen, así como la validación de las condiciones apropiadas del paciente, la preparación correcta, etc.;
- La validación de la muestra se ocupa de obtener el espécimen correcto en el momento adecuado del paciente correcto, el procesamiento y transporte correctos de la muestra, y el uso correcto de esa muestra para las mediciones analíticas;
- La validación técnica tiene que ver con obtener la respuesta correcta, que requiere el conocimiento de la calidad requerida para un examen, la validación de la precisión y la exactitud del proceso de medición, diseñando el procedimiento de Control Estadístico Interno de la Calidad correcto, e implementando los procedimientos de medición y control apropiadamente;
- La validación del paciente requiere que los resultados correctos sean reportados correctamente al registro indicado del paciente y considera si el resultado del examen es consistente con el conocimiento que se tenga del paciente (antecedentes), otros resultados de examen de ese paciente, dentro de la variación esperada del paciente individual y el grupo de población apropiada, así como la relación de los valores críticos o de alerta, y la consistencia con las poblaciones de pacientes;
- La validación clínica se ocupa de que el paciente reciba el tratamiento clínico correcto basado en los resultados de los exámenes y en los servicios del laboratorio. La validación clínica va más allá de lo que se considera normalmente como parte de la validación de los resultados de exámenes en E.U.A.

Este es un modelo europeo y demuestra que el paciente y la validación clínica han sido por largo tiempo una parte crítica en la validación de los exámenes de laboratorio. La validación clínica y de los pacientes son responsabilidades de los profesionales del laboratorio. Con el incremento en la carga de trabajo, se han desarrollado herramientas computarizadas y programas para estandarizar y facilitar esta “revisión clínica” o “Control de la Calidad Clínico”. La importancia del primer paso (validación administrativa) se vuelve clara en el contexto de la información necesaria para completar esta revisión y control clínico.

A continuación, consideremos los estudios más recientes y definitivos en las fuentes de errores del laboratorio (2). Los Drs. Plebani y Carraro han estudiado los errores del laboratorio por muchos años y son reconocidos como líderes en la realización de dichos estudios. Sus resultados documentan la distribución de errores como se muestra en la figura. Claramente hay más errores pre-analíticos (60%) y post-analíticos (25%) que analíticos (15%). Muchos científicos

del laboratorio clínico citan estas figuras para apoyar la idea de que los errores pre-analíticos y post-analíticos son más importantes que los errores analíticos y a menudo concluyen que la calidad analítica no es ya un asunto importante.

Sin embargo, una lectura más detallada del estudio muestra que de un total de 51.746 exámenes, hubo 393 resultados cuestionables, 160 de los cuales fueron confirmados como errores del laboratorio. De estos 160 errores, 46 causaron un cuidado inapropiado del paciente, y 24 de esos casos de cuidado inapropiado del paciente fueron causados por errores analíticos. Esto significa que más de la mitad de los casos (52%) de los errores que provocaron un cuidado inapropiado de la salud de los pacientes se deben a errores analíticos. ¡Los errores analíticos todavía son la fuente más grande e importante de errores que causan daño a los pacientes!

Necesitamos reconocer que la competencia principal de un laboratorio debe ser obtener los resultados correctos. Todas las fuentes de error en el proceso de examen deben ser controladas y monitoreadas cuidadosamente, pero se debe comenzar con los errores analíticos. Si no podemos, no queremos o no aseguramos la calidad analítica de nuestros resultados, entonces no deberíamos estar en el negocio de la provisión de servicios de exámenes de laboratorio.

¿Cuál es el punto?

Debe pensar de manera crítica sobre la calidad y reconocer que muchas de nuestras creencias actuales no están cimentadas en hechos, esto es, no están basadas en evidencias científicas. Estos mitos necesitan ser traídos a la luz si se va a mejorar el manejo técnico de los procesos analíticos.

¡Ese es el propósito de esta introducción! Debe evaluar críticamente muchas de las prácticas del manejo de la calidad que son aceptadas hoy en día en los laboratorios. Para comenzar, debe comprender cómo los requisitos de la calidad pueden ser definidos, cómo el desempeño del método debería ser medido experimentalmente, cómo los datos experimentales pueden ser analizados con estadísticas para estimar las características analíticas de desempeño, y cómo puede ser juzgada la aceptabilidad del procedimiento de examen.

Una vez que el desempeño de un método ha sido juzgado como aceptable (Validación Básica de Métodos), necesita seleccionar un procedimiento de Control Estadístico Interno de la Calidad que pueda detectar errores clínicamente significativos (Planificación de la Calidad o Diseño de la Calidad), hacer mediciones rutinarias del número necesario de controles e interpretar los resultados del control usando los criterios de decisión apropiados o las reglas de control (Prácticas Básicas de Control de la Calidad).

Referencias

1. Oosterhuis WP, Ulenkate HJLM, Goldschmidt HMJ. Evaluation of LabRespond, a new automated validation system for clinical Laboratory test results. *Clin Chem* 2001; 46:1811-1817.
2. Carraro P, Plabani M. Errors in a Stat Laboratory: Types and frequencies 10 years later. *Clin Chem* 2007; 53:1338-1342.

2: ¿Cómo gestiona usted a la calidad?

Los principios del Gestión Total de la Calidad (TQM por sus siglas en inglés) son adaptados para definir un marco de trabajo para el manejo de la calidad en un laboratorio al servicio de la salud. Este marco identifica cinco componentes fundamentales que son necesarios en un proceso de manejo de la calidad para proveer la mejora continua de la calidad.

Objetivos:

- Identificar los componentes fundamentales para el manejo de la calidad.
- Explicar el propósito de cada componente.
- Dar ejemplos de actividades asociadas con cada componente.
- Reconocer la fuerza conductora central en este proceso de manejo de la calidad.
- Describir el rol de la validación de métodos en este proceso de manejo de la calidad.
-

Materiales del capítulo:

- Gestión de la Calidad – La necesidad de procesos estándar y estándares para la calidad, por James O. Westgard PhD

Cosas qué hacer:

- Estudiar la lección.

Preguntas de Autoevaluación:

- ¿Cuáles son los 5 componentes necesarios para gestionar a la calidad?
- ¿Dónde se inserta la validación de métodos en este proceso de gestión de la calidad?
- ¿Por qué es importante la validación de métodos?
- ¿Qué es lo que guía o dirige un proceso de gestión de la calidad?

Gestión de la Calidad

La necesidad de procesos estándar y los estándares para la calidad

James O. Westgard, PhD

La calidad debe ser asegurada, no sobreentendida! Tal como se ilustra en el capítulo anterior, las ideas acerca del estado actual de la calidad en los exámenes de laboratorio y el cuidado de la salud podrían estar influenciadas por cuentos equívocos, teorías e hipótesis o mitos que no están sustentados por datos o hechos. ¡La calidad no ocurre espontáneamente! La calidad debe ser obtenida mediante procesos de trabajo que estén cuidadosamente planeados, ejecutados eficazmente, controlados óptimamente, medidos apropiadamente y continuamente mejorados, en otras palabras, mediante el manejo adecuado de la calidad. Esta lección enfatiza la necesidad de procesos estándar de laboratorio para proveer una calidad consistente, así como también estándares para la calidad para guiar el manejo de esos procesos.

Estándar, procesos estándar, estándares para la calidad

“Estándar” tiene varios significados en el laboratorio. En términos analíticos, una solución estándar provee un valor conocido para la calibración de un proceso de examen y representa pureza, veracidad y exactitud. En términos de gestión, un proceso estándar provee una manera consistente de realizar las cosas, por ejemplo, un proceso de examen estándar brinda un protocolo bien definido para realizar un examen de laboratorio. De la misma manera, un proceso estándar de validación de métodos brinda un modo regular y sistemático para evaluar el desempeño de un proceso de examen de laboratorio. (Nótese que la idea de proceso puede ser aplicada a cualquier actividad repetida, sea física o mental; tanto los procesos de trabajo físico como de toma de decisiones necesitan ser estandarizados o sistematizados para asegurar resultados consistentes y confiables).

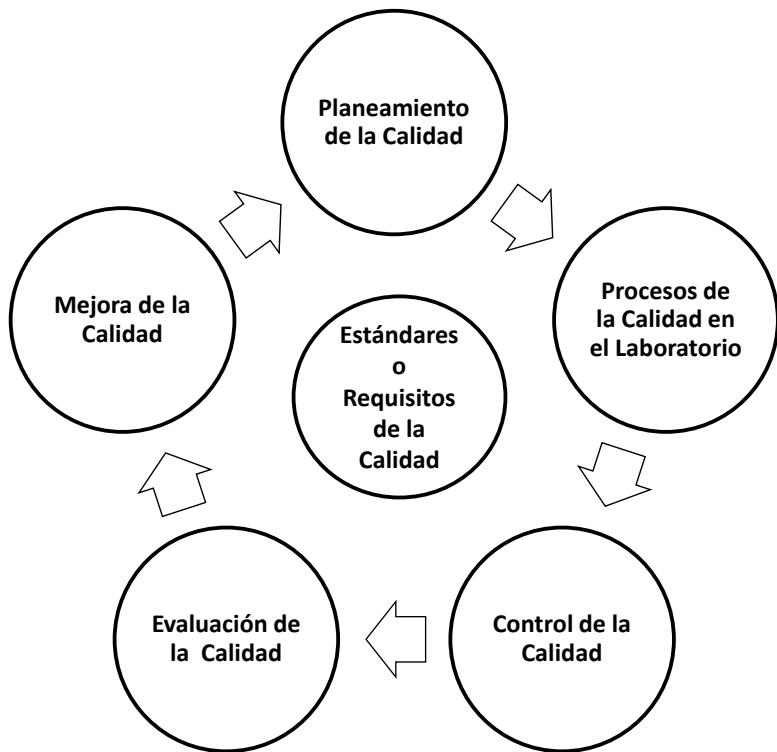
Otro término importante es estándar de la calidad, que es un criterio o declaración que describe el nivel aceptable de alguna característica. Para un examen analítico, debemos saber qué tan rápidamente necesitamos reportar el resultado de un examen, así como también qué tan cercano debe estar el resultado al valor correcto o verdadero. Un estándar para el tiempo requerido es más fácilmente entendido que un estándar para veracidad o exactitud.

Por ejemplo, es obvio que el tiempo requerido debería ser estimado en

unidades de tiempo, generalmente en minutos. Estas unidades son entendidas tanto por quien solicita el examen como por quien provee el servicio. Aquel que ordena el examen define el requisito en base al servicio médico que se ofrece. Ambas partes pueden medir si el desempeño observado satisface el requisito o no. La calidad analítica es más difícil ya que involucra conceptos técnicos tales como la imprecisión y la inexactitud, que no son siempre comprendidos por los laboratoristas y son ciertamente menos aún entendidos por los médicos que ordenan los exámenes o los pacientes, que son los últimos consumidores de los resultados de los exámenes. Los clientes y los consumidores de los servicios del laboratorio no pueden definir fácilmente a la calidad analítica que es requerida (al menos no en los términos analíticos deseados por el laboratorio) ni tampoco medir o evaluar la calidad analítica. El laboratorio, por lo tanto, debe asumir la completa responsabilidad del manejo de la calidad analítica de sus servicios.

Un proceso estándar para la gestión de la calidad

La gestión de la calidad debería ser un proceso estándar de laboratorio. Dicho proceso puede ser estructurado como se muestra en la siguiente figura. Comenzando por la parte superior y avanzando en el sentido de las agujas del reloj, PC, se refiere al planeamiento de la calidad, PCL, a los procesos de la calidad en el laboratorio, CC, al control de la calidad, EC, a la evaluación de la calidad, y MC, a la mejora de la calidad. En el centro, SC, significa estándares de la calidad, en otras palabras, estándares para la calidad requerida por el examen



o el servicio que se provee. Este marco de trabajo brinda un proceso de manejo de la calidad que funciona como un círculo de retroalimentación. PC, planea el mejor modo de realizar el trabajo, por ejemplo, la selección y evaluación de los métodos analíticos, el equipamiento, los reactivos y los procedimientos utilizados para llevar a cabo exámenes de laboratorio. El PCL establece los procesos de trabajo estándar para utilizar las políticas, procedimientos, protocolos y personal del laboratorio; el CC provee mediciones cuantitativas del desempeño del proceso usando técnicas estadísticas de control de procesos; la EC brinda mediciones más amplias acerca de qué tan bien se está realizando el trabajo, por ejemplo, la efectividad en los procedimientos de toma de muestras, el tiempo requerido para los servicios del laboratorio, los formatos apropiados para el reporte de resultados, etc. Cuando se detectan problemas, MC provee un mecanismo de solución de problemas para determinar sus causas, las cuales pueden luego ser eliminadas a través de PC, en cuyo caso, los procesos de examen son planeados nuevamente y se implementan nuevos y mejores modos de realizar el trabajo (es decir, cambios en el PCL). A través de este marco de trabajo, la mejora continua se construye dentro del proceso de gestión por medio de las “C”. El enfoque al cliente se logra al centrar estas “C” en los estándares de la calidad que representan los objetivos del laboratorio y los requisitos del cliente – información esencial para el manejo objetivo y cuantitativo del laboratorio.

El enfoque inicial del laboratorio es el establecimiento de procedimientos o procesos operativos estándar por medio de los cuales se realiza el trabajo. En particular, nos ocupamos de la selección, la validación y la implementación de métodos analíticos que brindarán consistentemente una calidad analítica confiable. Estas actividades son parte del PC y conducen al establecimiento de procesos de exámenes de calidad que son parte del PCL. El CC se aplica luego de que se establezcan los procesos de exámenes apropiados. De la misma manera, la EC y la MC son parte del monitoreo y mejora continua de los procesos operativos estándar.

Prioridades en el desarrollo de un proceso estándar de gestión de la calidad

Las prácticas tradicionales para la gestión de la calidad en el laboratorio siempre han comenzado con el establecimiento de métodos estándar para realizar exámenes de laboratorio (PCL), luego han enfatizado el control estadístico de la calidad para monitorear el desempeño analítico (CC), y más tarde se han expandido y ampliado las mediciones de la calidad para incluir el tiempo requerido, etc. (EC). Las prácticas de mejora de la calidad (MC) han sido introducidas como parte del Gestión Total de la Calidad (MTC) o Mejora Continua de la Calidad (MCC) [2,3] así como “*Seis Sigma*” [15].

Los laboratorios generalmente carecen de:

- Estándares de la Calidad (SC): metas definidas de la calidad, objetivos y requisitos de los clientes, y
- Planeamiento de la Calidad (PC): metodología y destrezas para construir la calidad deseada en los procesos de laboratorio.

Si estos requisitos de la calidad y los planeamientos para alcanzarlos no han sido especificados, la calidad no está siendo realmente gestionada. En cambio, lo que sucede, es lo que sucede. Es como en los dibujos animados donde Lucky Eddy y Hagar El Horrible están en un bar. Hagar le pregunta a Lucky Eddy qué va a beber y Lucky Eddy le contesta “Lo de siempre”. Hagar le pregunta “¿Qué es lo de siempre?” y Lucky Eddy le responde “Es lo que bebo siempre”. En laboratorios al servicio de la salud, a pesar de las supuestamente bien establecidas prácticas de gestión de PCL, CC, EC y MTC, lo que pasa con la calidad es generalmente lo que ha pasado antes. Para gestionar realmente a la calidad, los laboratorios deben definir la calidad que es requerida e implementar los procesos sistemáticos para validar el desempeño del método, seleccionar los procedimientos de control estadístico interno de la calidad apropiados y monitorear el desempeño del proceso de manera objetiva y en forma cuantitativa.

Una breve historia sobre los Estándares de la calidad

En realidad, hay una larga historia de debates sobre estándares de la calidad en la literatura de la química clínica. Comenzando en 1963, un químico clínico canadiense de nombre David Tonks, sugirió que los errores que son permitidos en un examen deberían ser relacionados con el ancho del intervalo de referencia (en ese momento llamado “rango normal”) [4]. Un examen que tenga un intervalo de referencia estrecho debería tener también alta precisión para distinguir si un paciente era “normal” o no. Tonks recomendó que el error permitido para un examen no debiera ser mayor de un cuarto del rango normal.

En 1968, el doctor Roy Barnett, un patólogo clínico, investigó la relación entre el desempeño a nivel de precisión de los métodos de laboratorio y el uso clínico y la interpretación de los resultados de los exámenes por parte de los médicos. En su famoso artículo acerca de “El significado médico de los Resultados del Laboratorio” [5], Barnett brindó lineamientos específicos para las desviaciones estándar permitidas médicaamente (S) para muchos de los exámenes de laboratorio de rutina.

Debido a que Tonks definió a la calidad deseada en la forma de un error permitido, y Barnett lo hizo en la forma de una desviación estándar permitida, hubo desacuerdo acerca del modo correcto de dar formato a las recomendaciones

para los estándares de la calidad. En 1976, el *College of American Pathologists* (CAP) convocó a una conferencia para debatir las metas analíticas. Como yo era un joven químico clínico que había escrito artículos sobre la evaluación de métodos y la necesidad de estándares de la calidad para juzgar la aceptabilidad de nuevos métodos, fui invitado a esta conferencia para exponer. Propuse el concepto de “error total” y el uso del “error total permitido” como el mejor formato para definir a la calidad, pero me encontré en el lado opuesto del pensamiento de los bien establecidos químicos y patólogos clínicos. Los debates de esta “Conferencia de Aspen” brindan algunos vistazos de los altercados y desacuerdos [6]. Éstos continuaron por muchos años e inhibieron la aplicación práctica de cualquier forma de estándar de la calidad en los laboratorios clínicos. Durante este tiempo, el concepto error total se consolidó y fue aceptado en los laboratorios clínicos. Finalmente en 1999, la Conferencia de Consenso de Estocolmo convocó a muchos expertos en el campo para “establecer especificaciones de la calidad analítica globales” [7]. El resultado fue una recomendación para una jerarquía de “especificaciones de la calidad” para priorizar la fuente preferida de información, como sigue:

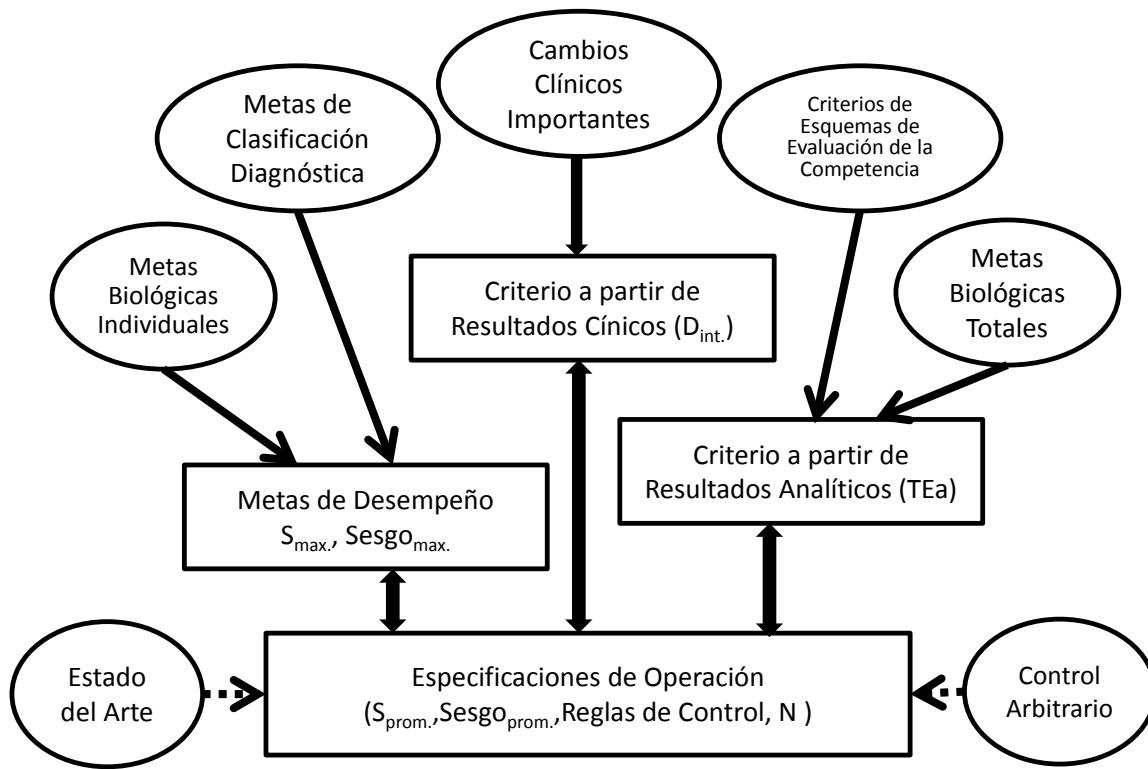
- “La evaluación del efecto del desempeño analítico de resultados clínicos en un marco clínico específico;
- La evaluación del efecto del desempeño analítico en las decisiones clínicas en general:
 - Información basada en componentes de variación biológica,
 - Información basada en el análisis de opiniones de clínicos;
- Recomendaciones profesionales publicadas:
 - De grupos expertos nacionales e internacionales;
 - De grupos expertos locales o individuos
- Metas de desempeño establecidas por
 - Entes regulatorios;
 - Organizadores de Programas de la Evaluación Externa de la Calidad (PEEC);
- Metas basadas en el estado actual de la técnica:
 - Tal como se demuestra en la información proveniente de la programas de evaluación externa de la calidad o de esquemas de evaluación de la competencia;
 - Como se encuentra en la publicaciones actuales de metodología”.

Cuando estén disponibles y cuando sean apropiado para el propósito deseado, se preferirán los modelos más altos en la jerarquía a los modelos de niveles más bajos” [7]. Por lo tanto, se prefieren los modelos específicos clínicos a los modelos clínicos generales, los cuales a su vez son preferidos por sobre las recomendaciones de grupos expertos, el criterio de Esquemas de Evaluación de la Competencia o programas de Evaluación Externa de la Calidad y los lineamientos del estado actual de la técnica (estado del arte). En la práctica, sin embargo, un laboratorio quizás debe satisfacer requisitos regulatorios, que entonces tienen prioridad sobre los modelos clínicos y las metas biológicas, aún si el criterio de los Esquemas de Evaluación de la Competencia y Programas de Evaluación Externa de la Calidad se encuentra por debajo en la jerarquía.

Una nota final: la terminología varía ampliamente a través del tiempo entre los científicos, de un texto a otro, entre profesiones, entre partes involucradas en regular a la calidad, entre países, entre continentes, en el mundo entero. Uno debe adaptarse a nombres y términos diferentes al leer, estudiar e investigar la literatura como guía. Yo tiendo a utilizar los términos “metas de la calidad” y “requisitos de la calidad” como sinónimos de “estándares de la calidad” y reservar el término *especificaciones* para la precisión y exactitud necesarias para alcanzar la meta como en “especificaciones operativas”. Pero muchos de mis artículos anteriores muestran el uso de “estándares de la calidad” como se refleja en este libro. Perdón por la confusión, pero es una parte integral en la historia de los estándares de la calidad.

Ponerse en marcha con los estándares de la calidad

Nuestro objetivo aquí no es que usted se transforme en un experto en el tema de estándares de la calidad, sino ayudarlo a que comience a aprender acerca de los estándares de la calidad y asegurarnos que usted conoce las diferentes fuentes de información que están convenientemente disponibles. No existe un solo y único modo de definir a la calidad necesaria para un examen de laboratorio, aun cuando hay feroces debates a favor y en contra de ciertos tipos de requisitos de la calidad. Se necesitan diversos tipos de estándares de la calidad para el manejo de la calidad en diferentes momentos del proceso, tales como criterios con orientación clínica que reflejan cambios clínicamente importantes en los resultados de los exámenes, criterios con orientación analítica, que describen el error analítico total permitido en los resultados de los exámenes, y especificaciones operativas analíticas que describen la imprecisión permitida, el sesgo permitido y el control estadístico interno de la calidad necesario para detectar errores clínicamente importantes en el proceso de examen.



La figura adjunta muestra las relaciones entre ciertos tipos de recomendaciones y diferentes tipos de criterios de la calidad. Comenzando por la parte superior de la figura, los cambios clínicamente importantes en los resultados de los exámenes pueden ser definidos por medio de lineamientos de tratamiento estándar (vías clínicas, lineamientos clínicos prácticos, etc.) para establecer criterios con fundamentos clínicos (o intervalos de decisión, $D_{int.}$). Dichos criterios clínicos pueden ser convertidos en especificaciones operativas de laboratorio para la imprecisión ($S_{prom.}$), inexactitud ($sesgo_{prom.}$), y Control Estadístico Interno de la Calidad (reglas de control, N) por un modelo clínico de planificación de la calidad [8] que toma en cuenta factores pre-analíticos, tales como variabilidad biológica individual o intra-individuo($S_{intra\ Ind.}$).

El lado derecho de la figura muestra cómo los criterios de Esquemas de Evaluación de la Competencia o Programas de Evaluación Externa de la Calidad definen los criterios con fundamentos analíticos en la forma de errores totales permitidos (TE_a), los cuales pueden ser convertidos en especificaciones operativas de laboratorio para la imprecisión ($S_{prom.}$), inexactitud ($sesgo_{prom.}$), y Control Estadístico Interno de la Calidad (reglas de control, N) por medio de un modelo de planificación de la calidad analítica [9]. Nótese que el error total permitido también puede ser establecido en base a metas biológicas totales que pueden estar basadas en datos poblacionales o individuales, por lo tanto la

base de datos extensiva de variabilidad biológica individual puede también ser usada en esta situación [10].

El lado izquierdo de la figura muestra cómo los criterios de desempeño para la imprecisión e inexactitud pueden ser definidos como metas analíticas separadas para la máxima imprecisión y sesgo que serían permitidos para el desempeño estable del método. Las especificaciones para la imprecisión máxima pueden ser derivadas considerando la variabilidad biológica intra-individuo [11]. El sesgo máximo permitido puede ser derivado de modelos de clasificación de diagnóstico [12]. Los laboratorios pueden utilizar estos criterios de desempeño individuales relacionando el desempeño del método observado con el valor máximo permitido, calculando el error de tamaño crítico que necesita ser detectado para mantener un desempeño satisfactorio y luego seleccionando procedimientos de Control Estadístico Interno de la Calidad apropiados por medio del uso de gráficos de función de potencia.

La parte inferior de la figura muestra qué pasa en la ausencia de estándares definidos de la calidad. El desempeño analítico del “estado de la técnica” (estado del arte) establece las especificaciones para la imprecisión y la inexactitud ya que los fabricantes fijan los objetivos de desempeño de sus productos en base al desempeño necesario para poder ser competitivos en el mercado. Se establece entonces un control arbitrario en lugar del control estadístico interno de la calidad necesario porque las prácticas de Control Estadístico Interno de la Calidad son establecidas en base a lineamientos de la práctica profesional, regulatorios o de acreditación.

Fuentes convenientes de estándares de la calidad

- Los criterios obtenidos a partir de Esquemas de Evaluación de la Competencia o Programas de Evaluación Externa de la Calidad ofrecen un listado de requisitos analíticos de la calidad para un desempeño aceptable que han sido definidos en CLIA (por sus siglas en inglés “Enmiendas para el Mejoramiento de los Laboratorios Clínicos”). Esta información está fácilmente disponible y puede ser utilizada al validar el desempeño de métodos analíticos.

Véase Apéndice 1 o <http://www.westgard.com/clia.htm>.

- También se encuentra disponible información acerca de los cambios clínicamente importantes en los resultados de exámenes, aunque dicha información debe ser interpretada cuidadosamente si va a ser utilizada para validar el desempeño de métodos analíticos.

Véase <http://www.westgard.com/clinical.htm>

- Puede consultarse una base extensiva de datos en línea que sintetizan todos los estudios disponibles de variabilidad biológica y ofrecen recomendaciones para las especificaciones de precisión, sesgo y error total. Esta base de datos es una actualización del trabajo original publicado por la doctora Carmen Ricos y sus colegas españoles [13]. Para el debate concreto de variabilidad biológica, véase el libro del doctor Callum Fraser [14].

Véase <http://www.westgard.com/biodatabase1.htm>

- También puede consultar recomendaciones europeas anteriores para metas biológicas para impresión e inexactitud, así como errores totales biológicos permitidos basados en variabilidad biológica individual o intra-individuo.

Véase <http://www.westgard.com/europe.htm>

¿Cuál es el punto?

La validación de métodos debería ser un proceso estándar en su laboratorio. Debería ser parte de sus procedimientos operativos estándar para realizar el trabajo en su laboratorio. Eso significa tener un protocolo que defina un conjunto estándares de procedimientos experimentales, modos estándar de recolectar y analizar la información y una manera estándar de juzgar la aceptabilidad de un nuevo método. ¡Y esa decisión de aceptabilidad va a depender de la definición de un estándar de la calidad que brinde una declaración objetiva sobre que tan bueno debe ser un examen! Un punto de partida práctico para los estándares de la calidad podría ser seleccionar algún requisito de la calidad regulatorio que defina “criterios de desempeño aceptable” como los requisitos de CLIA para los Esquemas de Evaluación de la Competencia. Pero el número de exámenes en el listado de CLIA es limitado y será necesario consultar otras fuentes y recomendaciones, tales como aquellas basadas en la variabilidad biológica intra-individuo y entre-individuos y, cuando se hallen disponibles, requisitos basados en el uso clínico deseado de los exámenes de laboratorio.

Estas lecciones acerca de la validación de métodos le enseñarán un proceso estándar, aprenderá cómo realizar los experimentos apropiados y recoger la cantidad necesaria de información y cómo analizar esa información de manera analítica y gráfica para interpretar críticamente los resultados experimentales. Una decisión objetiva sobre la aceptabilidad de cualquier método dependerá de la capacidad de definir un estándar de la calidad para ese examen. Esa es quizás la evaluación más crítica y difícil de hacer, pero una vez que se tiene esa información, la decisión acerca del desempeño del método es simple y directa.

Tendencias y orientaciones

El foco aquí está puesto en el proceso básico para el manejo de la calidad y la información actualmente disponible sobre los estándares de la calidad. Pero por supuesto, las cosas cambiarán y evolucionarán y es de esperar que haya nuevas tendencias y orientaciones.

En primer lugar, los criterios CLIA para el desempeño aceptable fueron formulados originalmente en 1992 y no han sido aún actualizados. Se espera que el listado CLIA de analitos regulados cambie, tanto los analitos en el listado como los valores numéricos para los criterios de desempeño aceptable. Las regulaciones CLIA incluyen una disposición para actualizar estos criterios pero CDC/CMS no ha hecho cambios en los primeros quince años. Mientras tanto, los exámenes en los laboratorios han avanzado, pero no hay criterios CLIA para exámenes importantes como HbA1c, cTn y PSA y nuevos exámenes como hsCRP. Es obvio que se necesitan cambios.

En segundo lugar, hay que tener en cuenta que siempre hay nuevas propuestas. La gestión tiende a “estar a la moda”, siempre buscando *Nuevas* formas que traerán *Nuevos* beneficios. Lo mismo aplica para la gestión de la calidad. Hoy en día los laboratorios pueden tener programas para La Gestión Total de la Calidad, Seis Sigma, Lean, Seguridad del Paciente, Análisis de Riesgos e Indicadores de la Calidad. Pueden seguir lineamientos de CLSI (Instituto de Estándares de Laboratorios Clínicos, por sus siglas en inglés) y de ISO (Organización Internacional de Estándares, por sus siglas en inglés). Están sujetos a inspecciones de una variedad de acreditadores, tales como CMS (Centros para Servicios de Cuidado y Ayuda Médica), CAP (Colegio de Patólogos Americanos), JC (Comisión Conjunta para la Acreditación de Organizaciones al Cuidado de la Salud), y COLA (originalmente llamada la Comisión para la Acreditación de Laboratorios). Desde mi punto de vista, todas estas propuestas, programas y lineamientos son herramientas que deberían ser integradas al proceso básico de gestión de la calidad descripto aquí, en vez de ser sustituidas por este proceso básico.

Referencias

1. Westgard JO, Barry PL. Total quality control: Evolution of quality management systems. *Laboratory Medicine* 1989;20:377-384.
2. Westgard JO, Barry PL. Beyond quality assurance: Committing to quality improvement. *Laboratory Medicine* 1989;20:241-247.
3. Westgard JO, Burnett RW, Bowers GN. Quality management science in clinical chemistry: a dynamic framework for continuous improvement of quality. *Clin Chem* 1990;36:1712-1716.
4. Tonks DB. A study of the accuracy and precision of clinical laboratory determinations in 170 Canadian laboratories. *Clin Chem* 1963;9:217-233.
5. Barnett RN. Medical significance of laboratory results. *Am J Clin Path* 1968;50:671-676.
6. 1976 Aspen Conference on Analytical Goals in Clinical Chemistry.
7. Hyltoft Petersen P, Fraser CG, Kallner A, Kenny D. Strategies to set global analytical quality specifications in laboratory medicine. *Scand J Clin Lab Invest* 1999;59:No.7(Nov).
8. Westgard JO, Hyltoft Petersen P, Wiebe DA. Laboratory process specifications for assuring quality in the U.S. National Cholesterol.
9. Westgard JO, Wiebe DA. Cholesterol operational process specifications for assuring the quality required by CLIA proficiency testing. *Clin Chem* 1991;37:1938-1944.
10. Hyltoft-Petersen P, Ricos C, Stockl D, Libeer J-C, Baadenhuijsen H, Fraser CG, Thienpont L. Proposed guidelines for the internal quality control of analytical results in the medical laboratory. *Eur J Clin Chem Clin Biochem* 1996;34:983-999.
11. Fraser CG, Hyltoft Petersen P, Ricos C, Haeckel R. Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Clin Biochem* 1992;30:311-317.
12. Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin Chem* 1993;39:1514-1518.
13. Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, Minchinela J, Perich C, Simon M. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999;59:491-500.

14. Fraser C. Biologic Variation: From principles to practice. AACC Press, Washington DC, 2001.
15. Westgard JO. Six Sigma Quality Design and Control: Desirable precision and requisite QC for laboratory testing processes. Second Edition. Madison, WI:Westgard QC, Inc., 2007.

3: ¿Cuál es el propósito de un estudio de validación de métodos?

El Dr. Westgard revela el significado secreto más interno, escondido y profundo de la validación de métodos. El conocimiento de este “significado” debería hacer todo el proceso de validación de métodos más racional y entendible.

Objetivos:

- Comprender qué estudios de validación de métodos se deben estudiar.
- Reconocer las deficiencias potenciales de la estadística en los estudios de validación de métodos.
- Identificar los diferentes tipos de errores analíticos que necesitan ser evaluados.

Materiales del capítulo:

- VM – **El significado Secreto más Interno, Escondido y Profundo,** por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar la lección.
- Revisar un artículo de validación de métodos de la literatura científica.

Preguntas de Autoevaluación:

- ¿Cuáles son los dos tipos principales de errores analíticos?
- ¿Qué se entiende por Error Total?
- ¿Cómo se relaciona el Error Total con los tipos de errores básicos?
- ¿Cómo describe la literatura los errores de un método?
- ¿Qué estadísticos se emplean en un informe de validación de métodos?
- ¿Cómo se vinculan las conclusiones del informe con los errores del método?

Validación de Métodos: El Significado Secreto más Interno, Escondido y Profundo

James O. Westgard, PhD

En el primer año de escuela, tuve un profesor de inglés que me enseñó algo que nunca he olvidado. Él siempre preguntaba, “¿cuál es el significado secreto más interno, escondido y profundo de lo que están escribiendo?” En otras palabras, ¿qué estás tratando realmente de alcanzar? Será mejor que lo descubras si estás esperando que alguien más lo comprenda. En verdad, puedes escribir acerca de algo, pero tienes que ser realmente claro en lo que quieras lograr, si esto no es así, le verdadero si propósito y significado permanecerá oculto.

La verdadera sorpresa llegó en mi primer trabajo como químico clínico cuando comencé a evaluar el desempeño de un nuevo equipo de química de ensayos múltiples. Estudié toda la literatura científica existente que proveen los lineamientos para realizar estudios de validación de métodos (VM), pero no estaba en absoluto claro cómo decir si un método era o no aceptable. ¡Nadie decía el secreto! Y ese secreto es de suprema importancia para evaluar un método apropiadamente. Claro, puedes recolectar algunos datos, calcular algunas estadísticas y proveer algo de papeleo en una carpeta para mostrarlo a un inspector de laboratorios, pero ¿es esta la razón por la que realmente lo haces?

En tanto que no afirmaré que mi profesor de inglés me hizo un mejor escritor (tampoco lo pueden culpar ustedes), me hizo un mejor científico ayudándome a buscar el significado más profundo y el propósito real de lo que hago. ¿Cuál es el propósito verdadero de la validación de métodos? ¿Cuál es el problema que estamos tratando de resolver? ¿Brinda la práctica actual soluciones correctas? ¿Existe una forma mejor de hacer esto? ¿Cómo sabes cuál es la forma correcta de validar el desempeño de un método?

El Secreto Revelado

Aquí está el significado secreto más interno, escondido y profundo de la validación de métodos (y no tienes que leer más allá para entender el mensaje): EVALUACIÓN DE ERRORES. Uno quiere estimar qué tanto error podría estar presente en el resultado de un examen producido por un método en su laboratorio. Con esta información, entonces usted quiere asegurar que la magnitud del error no afectará la interpretación del resultado del examen y comprometerá el cuidado del paciente. Si los errores observados son tan grandes que causarán una interpretación incorrecta, el método no es aceptable. Para que sea aceptable,

los errores observados necesitan ser pequeños en relación a los cambios que causarán un cambio en la interpretación de un resultado de un examen.

Un enfoque en los errores analíticos es la clave para todo proceso de validación de métodos. ¿Qué tipos de errores analíticos pueden ocurrir en un método de laboratorio? ¿Qué experimentos pueden proveer datos acerca de estos errores? ¿Cuál es el mejor modo de llevar a cabo esos experimentos para evaluar los errores? ¿Cuántos datos necesitan ser recolectados para obtener buenas estimaciones del error evaluado? ¿Qué estadística estima mejor el tamaño de estos errores a partir de los datos experimentales? ¿Qué tamaño de error es permitido sin afectar la interpretación de un examen y comprometer el cuidado del paciente?

La validación de métodos guarda estrecha relación con la evaluación de errores... ¡ese es el secreto!

Una prueba rápida

El coeficiente de correlación es un estadístico que casi siempre se calcula y se informa para describir los resultados de un estudio de comparación de métodos. Un valor de 1.000 indica una correlación perfecta entre los resultados de dos métodos. Otras estadísticos (tales como pendiente, ordenada al origen y desviación estándar de las diferencias) también pueden ser calculados a partir de los mismos datos para estimar el tamaño de error presente entre los dos métodos. ¿Cuáles son más útiles?

Considere la siguiente situación. Aquí hay un nuevo método de colesterol donde los resultados de un estudio de comparación de métodos dieron un coeficiente de correlación de 0.999, que es muy cercano al valor ideal de 1.000. Suena bastante bien, ¿no es así? ¿Qué tan cercanos son los resultados entre los dos métodos? ¿Es aceptable el nuevo método? Permitame darle un poco de información adicional.

Aquí está el gráfico de los resultados del examen del nuevo método frente a los del método de comparación. Nótese primero que el coeficiente de correlación muestra que los resultados son cercanos a la mejor línea de ajuste entre los métodos; no se muestra que los valores obtenidos a partir del método de prueba sean iguales a los obtenidos por el mé-



Resultados del Método de Comparación
Los resultados de un estudio de comparación, donde los valores del método "de prueba" o nuevo son graficados en el eje de las ordenadas y los valores de comparación en el eje de las abscisas

todo de comparación. El gráfico muestra que casi todos los valores del nuevo método son sistemáticamente más altos por 15 mg/dL. ¿Esta información de que hay un error sistemático de 15 mg/dL le ayuda a tomar una decisión sobre la aceptabilidad del nuevo método? No se ve tan bien ahora, ¿verdad? Siendo el error de 15 mg/dL puede ser limitada la utilidad de los resultados producidos por el nuevo método.

Como profesionales del laboratorio comprendemos intuitivamente los errores y tenemos un sentido de cómo éstos podrían afectar la interpretación de los resultados de los exámenes y el cuidado de los pacientes. ¡No tenemos el mismo sentido respecto a la estadística! Es por esto que las estadísticas deben ser usadas para estimar los errores que son significativos para nosotros [1] - este es un segundo secreto importante y se verá con mayor detalle más adelante.

De este sencillo ejemplo, usted podrá reconocer la dificultad en la interpretación del coeficiente de correlación, ya que no brinda por si solo una estimación útil de los errores analíticos. La información acerca del *tamaño* de los errores analíticos es más útil a la hora de juzgar el desempeño de un método [2]. El hecho de que el coeficiente de correlación se calcula comúnmente no lo hace útil. ¡Sólo muestra que la gente no sabe el secreto de la validación de métodos!

Errores analíticos

Enfoquémonos en los errores analíticos y asegurémonos de que tenemos un entendimiento general de los diferentes tipos de errores que necesitan ser estimados. Aquí hay una lista de términos que podrías necesitar entender: error aleatorio o imprecisión, error sistemático o inexactitud, error constante, error proporcional y error total.

Error aleatorio , EA, o Imprecisión

El error aleatorio es descrito como un error que puede ser tanto positivo como negativo, cuya dirección y magnitud exacta no pueden ser predichas, como se muestra en la figura a la derecha, por la distribución de los resultados de las mediciones repetidas efectuadas sobre un espécimen único.

La imprecisión es usualmente cuantificada calculando la desviación estándar (S) de los resultados de un conjunto de mediciones repetidas.



Error aleatorio (EA) o Imprecisión, como se muestra por la distribución de los valores de examen.

La S usualmente se incrementa conforme se incrementa la concentración, por lo que es generalmente útil calcular el coeficiente de variación (CV) para expresar la S como un porcentaje de la concentración media del estudio de repetibilidad. El tamaño máximo de un error aleatorio es expresado comúnmente como un estimado de 2 S ó 3 S para ayudar a comprender el tamaño potencial del error que podría ocurrir.

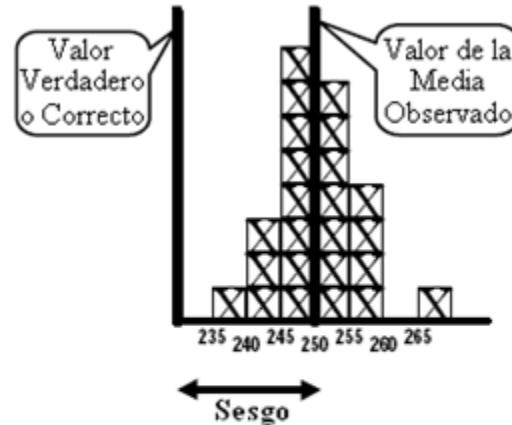
Error Sistemático, ES, o Inexactitud

Un error sistemático es un error que siempre va en una dirección, como se muestra en la figura a la derecha, donde un cambio sistemático desplaza la media de la distribución de su valor original. En contraste a los errores aleatorios que pueden ser tanto negativos como positivos y cuya dirección no puede ser predicha, los errores sistemáticos son en una sola dirección y causan que todos los resultados de exámenes sean altos o bajos.

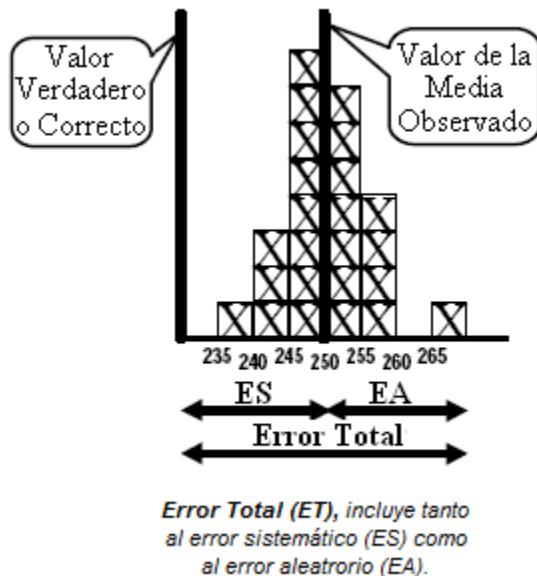
Qué tan altos o qué tan bajos puede ser descrito por el sesgo, que es calculado como la diferencia promedio, o la diferencia entre los promedios de los valores obtenidos a partir del método de “examen” y los valores obtenidos a partir de un método de “comparación” en un experimento de compración de métodos. Alternativamente, la diferencia sistemática esperada puede ser predicha a partir de la ecuación de la línea que mejor se ajusta a la proyección gráfica de los valores obtenidos a partir del método de examen en el eje de las ordenadas “Y” contra los valores obtenidos a partir del método de comparación en el eje de las abscisas “X”. El ES puede mantenerse constante en un rango de concentraciones, en cuyo caso puede ser llamado también error sistemático constante, o puede variar al cambiar la concentración, en cuyo caso se le puede llamar error sistemático proporcional.

Error Total, ET

El error total es el efecto neto o combinado de los errores aleatorios y sistemáticos, como se muestra en la figura abajo. Representa el “peor caso” de una situación, o sólo qué tan erróneo puede ser el resultado de un examen debido a ambos errores, aleatorio y sistemático. Dado que los laboratorios hacen típicamente sólo una medición por cada examen, esa medición puede tener error producto del ES esperado, o sesgo, más 2 ó 3 S, dependiendo de cómo se cuantifique el efecto del EA.



Error Sistemático (ES) o Inexactitud, como se muestra por el cambio o sesgo entre el valor de la media un valor correcto.



Mientras que en el laboratorio nos gusta pensar en la imprecisión y la inexactitud como errores separados, el médico y el paciente experimentan el efecto total de los dos, o sea el error total. El error total provee una medida orientada al cliente o consumidor del desempeño del examen, lo que lo hace el parámetro más importante al juzgar la aceptabilidad de los errores analíticos.

¿Cuál es el punto?

¡Debe entender el “por qué” de la validación de métodos para entender “cómo” debería ser efectuada la validación de métodos! El “por qué” define el propósito, que es determinar la cantidad de error que podría ocurrir con un método. El “cómo” define los protocolos experimentales y los procedimientos de análisis de datos que proveen los estimados de los errores que podrían ocurrir con el nuevo método. ¡La validación de métodos guarda estrecha relación con los errores!

Las regulaciones de laboratorio en los E.U.A. requieren que el desempeño del método para cualquier nuevo método sea “verificado” previo a reportar los resultados de los exámenes de los pacientes. **Bajo la Regla Final de CLIA, los laboratorios deben verificar el rango reportable, la precisión, la exactitud y los intervalos de referencia para todos los métodos non-wavived implementados después del 24 de abril de 2003.** Para los métodos desarrollados en el propio laboratorio o modificados por el laboratorio, también deben ser verificadas las características adicionales de sensibilidad analítica (límite de detección) y de especificidad analítica (interferencia, recuperación).

También son apropiados estudios de intervalos de referencia más extensivos. La responsabilidad de la verificación o validación de métodos, por lo tanto, aún reside en cada laboratorio. Mientras que los fabricantes usualmente recolectarán datos de validación de métodos durante la instalación de nuevos sistemas analíticos, el laboratorio aún es responsable de ver que se hayan recolectado los datos adecuados y que estos datos muestren que los nuevos métodos brindan un desempeño aceptable en el laboratorio.

Tendencias y Direcciones

Los esfuerzos para proveer estándares mundiales de prácticas de laboratorio van a cambiar los términos y conceptos que usamos. ISO, la Organización de Estándares Internacionales (*International Standards Organization*), provee lineamientos específicos para los laboratorios dedicados al cuidado de la salud en su documento **15189 – Laboratorios de Análisis Clínicos – Requisitos particulares para la calidad y la competencia** [3]. Cubriremos éste tema con mayor detalle en el capítulo 21. Por ahora, mencionemos brevemente lo que evoca el enfoque de ISO:

“Cuando se describen el desempeño de los procedimientos y la fiabilidad de sus resultados, debe usarse la terminología de ISO. Los resultados deben ser comparables universalmente y esto requiere trazabilidad metrológica, la incertidumbre concomitante que indica fiabilidad debería ser obtenida de manera universal y transparente, y debería ser combinable”. [4]

Los conceptos y la terminología preferidos por la ISO son “veracidad” e “incertidumbre”. La **veracidad** es utilizada para describir la “proximidad entre la medida un número de valores infinitos medidos y un valor de referencia”. Esto es equivalente a los términos sesgo y error sistemático en este capítulo. La **incertidumbre de medición** es usada por la ISO para describir un “parámetro no negativo que caracteriza la dispersión de los valores atribuidos a un mensurando a partir de la información que se utiliza”, donde el mesurando se refiere al analito o al examen particular. La incertidumbre describe un rango de valores que corresponde a un resultado dado, por ejemplo, un resultado de 200 podría tener una “incertidumbre estándar” (S , CV) de 4 unidades o 2%, indicando que un valor de 200 representa una “incertidumbre expandida” de 192 a 208 unidades (intervalo de confianza de 95% o $2 S$). Este concepto suena y luce similar a la precisión, pero el estimado de la incertidumbre también incorpora cualquier sesgo o veracidad; así, es de hecho más cercano a la idea de error total.

Estos conceptos de veracidad e incertidumbre vienen del mundo de la metrología, donde se provee a los clientes con productos que tienen valores “*target*” ((blanco, verdaderos o diana) asignados junto con la incertidumbre asociada

a la estimación de ese valor “*target*” ((blanco, verdadero o diana). El enfoque de la ISO espera que los clientes sepan el significado de incertidumbre. El mundo de los laboratorios clínicos es diferente. Los clientes, médicos y pacientes, no necesariamente poseen el conocimiento sobre la ciencia de las mediciones y la incertidumbre en los resultados de los exámenes. Sería mejor si los laboratorios controlaran sus métodos analíticos para verificar el cumplimiento de la calidad clínica pretendida de los resultados, pero en ausencia de esto, será necesario informar al cliente de la “duda” real de los resultados reportados. La incertidumbre de medición es casi seguramente una parte de sus responsabilidades futuras.

Referencias

1. Westgard JO, Hunt MR. Use and interpretation of common statistical test in method comparison studies. *Clin Chem* 1973;19:49-57.
2. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974;20:825-33.
3. ISO/FDIS 15789 Medical Laboratories – Particular requirements for quality and competence. 2003. International Organization for Standards, Geneva Switz. (Note that 2° edition of 15189 was published in 2007 and 3° edition is expected to be released by the end of 2008.)
4. Dybkaer R. Setting quality specifications for the future with newer approaches to defining uncertainty in laboratory medicine. *Scand J Clin Lab Invest* 1999;37:579-584.

4: ¿Cuáles son los requisitos regulatorios para la validación de métodos?

La Dra. Sharon S. Ehrmeyer revisa los lineamientos de CLIA (*Clinical Laboratory Improvement Amendments*) del gobierno de E.U.A. que relacionan los requisitos de validación de métodos con la complejidad de los exámenes de laboratorio. También describe los requisitos de validación de métodos de las organizaciones profesionales de acreditación tales como el *College of American Pathologists* (CAP) y la *Joint Commission* (JC).

Objetivos:

- Clasificar los exámenes de laboratorio en base a las guías de complejidad de CLIA.
- Relacionar la clasificación de la complejidad de un examen con los estudios experimentales requeridos para validar o verificar las características de desempeño de un método de laboratorio.
- Comparar los requisitos regulatorios y de acreditación para la validación de métodos.

Materiales del capítulo:

- VM — Las regulaciones, por Sharon S. Ehrmeyer, PhD

Cosas qué hacer:

- Estudiar la lección.
- Acceder a la página web de CLIA para encontrar la última clasificación de exámenes: <http://www.cms.gov/clia/>.

Preguntas de Autoevaluación:

- ¿Cuál es la clasificación de los procedimientos de examen considerando su complejidad que es empleada por CLIA para establecer los requisitos para la validación de métodos?
- ¿Cuáles son los requisitos para la validación de métodos *waived*?
- ¿Cuáles son los requisitos para la validación de métodos *non-waived* aprobados por la FDA?
- ¿Cuáles son los requisitos para la validación de métodos *non-waived* no aprobados por la FDA o modificados por el laboratorio?

Validación de Métodos: Las regulaciones

Sharon S. Ehrmeyer, PhD

Hasta que la Regla Final de CLIA fue publicada, muchos laboratorios escapaban a los requisitos de validación del desempeño de sus métodos. Las reglas del 28 de febrero de 1992 [1] brindaron una excepción para los laboratorios que llevaban a cabo exámenes moderadamente complejos usando métodos y sistemas analíticos no modificados aprobados por la FDA. Estos laboratorios podían aceptar las especificaciones de desempeño del fabricante en lugar de realizar estudios de validación de métodos ellos mismos. Todo esto cambió el 24 de enero de 2003 con la emisión de la Regla Final de CLIA [2]. Ahora TODOS los métodos de examen *non-waived* (de complejidad moderada y alta) requieren una validación de desempeño realizada en el propio laboratorio. Los CMS (*Centers for Medicare & Medicaid Services*) otorgaron un periodo “educacional” durante el cual no se entregaron citaciones, pero desde el 1 de enero de 2008, los requisitos de validación de métodos deben ser cumplidos o el laboratorio será citado por deficiencias en esta área.

Si bien CLIA establece los requisitos *mínimos* de examen, los laboratorios pueden escoger entre cumplir los requisitos (los cuales podrían ser más estrictos) de organizaciones de acreditación con estatus CLIA, tales como la *Joint Commission* (JC) [3], *College of American Pathologists* (CAP) [4], y COLA (anteriormente la *Commission on Office Laboratory Accreditation*) [5]. También hay organizaciones de laboratorios estatales de salud en Washington y Nueva York, las cuales están aprobadas por el gobierno (tienen estatus de exentos) y que imponen requisitos específicos. Además, siempre es importante mantener en mente aquellas prácticas que se esperaría formaran parte de la responsabilidad profesional y buenas prácticas del laboratorio.

Sistemas de la Calidad y los Procesos Totales de Examen

La Regla Final de CLIA se enfoca en los “sistemas de la calidad”, un término nuevo en esta revisión. Los CMS identifican estos componentes clave en un sistema de la calidad:

- Sistemas generales del laboratorio
- Sistemas pre-analíticos

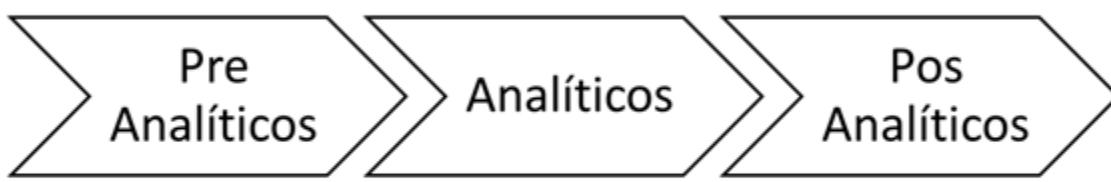
- Sistemas analíticos
- Sistemas post-analíticos

Todos estos componentes son importantes para asegurar la calidad de los servicios de examen de los laboratorios. Todos deben ser controlados, monitoreados y mejorados. En la Regla Final de CLIA, el proceso de asegurar la calidad de los servicios de examen de los laboratorios se llama evaluación de la calidad, más que aseguramiento de la calidad o gestión de la calidad, como se llamó en revisiones anteriores.

Independientemente de la terminología escogida, el intento de la Regla Final de CLIA es promover el desarrollo, la implementación, la entrega, el monitoreo y la mejora de los servicios de laboratorios de alta calidad. Esto requiere de un sistema de gestión de la calidad efectivo. En este contexto, *Sistema de la Calidad* se refiere a la organización, la estructura, los recursos, los procesos y los procedimientos necesarios para implementar la gestión de la calidad. Proceso significa una o más fuentes interrelacionadas y/o actividades que transforman entradas en salidas. *Procedimiento* representa una forma específica de realizar una actividad.

La Regla Final de CLIA está organizada alrededor del flujo de un espécimen a través del laboratorio. A este camino se le da el término de “proceso total de examen” e incluye los procesos y procedimientos pre-analíticos, analíticos y post-analíticos, como se muestra en la figura abajo.

•----- Proceso Total de Examen -----•



- | | | |
|----------------------------|--------------------------------|------------------------------------|
| •Preparación del paciente | •Fraccionamiento de la muestra | •Informe del resultado |
| •Toma de muestra | •Configuración del equipo | •Envío del informe |
| •Preparación de la muestra | •Calibración del ensayo | •Recepción del informe |
| •Transporte de la muestra | •Control de la Calidad | •Revisión de los resultados |
| •Pedido del medico | •Procesamiento de la muestra | •Acción a partir de los resultados |

El término *evaluación de la calidad* es usado para abarcar ampliamente las prácticas de medición y monitoreo de procesos y procedimientos, incluyendo validación y verificación, así como términos relacionados tales como evaluación, auditoría, etc. En nuestro enfoque del proceso analítico, la evaluación de la calidad debería involucrar metodología científica y análisis (estadística) de datos

para proveer una comparación objetiva del desempeño medido u observado con las metas y los requisitos de la calidad definidos.

Requisitos generales para el control de la calidad en los exámenes de laboratorio

Subpart K, Sistemas de la Calidad para Exámenes Nonwaived, comienza con los siguientes requisitos generales:

§493.1200 (a) Cada laboratorio que lleva a cabo exámenes *non-waived* debe establecer y mantener políticas y procedimientos escritos que implementen y monitorean sistemas de la calidad para todas las fases del proceso total de examen (esto es, pre-analítica, analítica y post-analítica), así como sistemas generales del laboratorio. (b) Los sistemas de la calidad de los laboratorios deben incluir un componente de evaluación de la calidad que asegure la mejora continua del desempeño del laboratorio y los servicios a través de un monitoreo en curso que identifique, evalúe y resuelva problemas. (c) Los diversos componentes del sistema de la calidad del laboratorio son usados para alcanzar los requisitos en esta parte y deben ser apropiados para las especialidades y subespecialidades de los exámenes que realiza el laboratorio, los servicios que ofrece y los clientes a quienes sirve.

§493.1250, Sistemas Analíticos, requiere a los laboratorios alcanzar los requisitos para lo siguiente:

§493.1251 Manual de procedimientos

§493.1252 Sistemas de examen, equipos, instrumentos, reactivos, materiales y suministros

§493.1253 Establecimiento y verificación de las especificaciones de desempeño del método

§493.1254 Chequeo del mantenimiento y la función

§493.1255 Procedimientos de calibración y de verificación de la calibración

§493.1256 Procedimientos de control

§493.1261-1278 Requisitos de especialidad y subespecialidad

§493.1281 Comparación de los resultados de examen

§493.1282 Acciones correctivas

§493.1283 Registro del examen

§493.1289 Evaluación de sistemas analíticos

En esta discusión, consideramos esos requisitos para el “establecimiento y verificación de las especificaciones del desempeño del método”, esto es, validación de métodos, más la comparación de los resultados de examen y la verificación de la calibración.

Validación de Métodos y Complejidad de Examen

Las recomendaciones que deben ser seguidas dependen de la “complejidad” del examen. La Regla Final de CLIA combinó complejidades moderadas y altas en una clasificación *Non-Waived* que gobierna la validación de métodos. La información acerca de la clasificación de los métodos de examen específicos está disponible en el website de CMS en la dirección:

<http://www.cms.gov/clia/>

Exámenes Waived. Los requisitos mínimos para los exámenes *waived* (baja complejidad) bajo CLIA es seguir las instrucciones de los fabricantes. No hay requisitos para la validación de métodos.

Exámenes Non-Waived Aprobados por la FDA. La mayoría de los exámenes realizados en los laboratorios actualmente caen en esta categoría, por lo que la mayoría de los estudios de validación de métodos deberían seguir las recomendaciones de la sección §493.1253:

§493.1253 Establecimiento y verificación de especificaciones de desempeño.

(a) Aplicabilidad. No se requiere a los laboratorios verificar o establecer especificaciones de desempeño para ningún sistema de examen usado por el laboratorio antes del 24 de abril de 2003.

(b)(1) Verificación de las especificaciones de desempeño. Cada laboratorio que introduce un sistema de examen no modificado, aprobado o autorizado por la FDA, debe seguir lo siguiente antes de reportar los resultados de exámenes de los pacientes:

(i) Demostrar que puede obtener—especificaciones de desempeño comparables a aquellas establecidas por el fabricante para las siguientes características de desempeño:

(A) Exactitud.

(B) Precisión.

(C) Rango reportable de los resultados de examen del sistema de examen.

- (ii) Verificar que los límites de referencia del fabricante (valores normales) son apropiados para la población de pacientes del laboratorio.

Esto generalmente significaría llevar a cabo cuatro experimentos:

- El experimento de comparación de métodos para estimar la inexactitud o sesgo;
- El experimento de replicación para estimar la imprecisión;
- El experimento de linealidad para determinar el rango reportable;
- La recolección de valores de referencia para verificar el intervalo de referencia [alternativamente, el director del laboratorio clínico puede documentar que los intervalos del fabricante o los intervalos del libro de texto son apropiados para los clientes a quienes sirve.]

Exámenes Non-Waived Modificados o Desarrollados En Casa. Los laboratorios que modifican exámenes de complejidad moderada o usan exámenes desarrollados en casa o exámenes clasificados como de alta complejidad bajo CLIA deben seguir requisitos de validación de métodos más estrictos, como sigue:

§493.1253 (b)(2) Cada laboratorio que modifica un sistema de examen aprobado o autorizado por la FDA, o que introduce un sistema de examen no sujeto a aprobación o autorización de la FDA (incluyendo métodos desarrollados en casa y métodos estandarizados tales como procedimientos de libros de texto, tinción Gram, o preparaciones de hidróxido de potasio), o usa sistemas de examen en los que las especificaciones de desempeño no son provistas por el fabricante, debe, antes de reportar los resultados de los pacientes, establecer para cada sistema de examen las especificaciones de desempeño para las características de desempeño siguientes, cuando sean aplicables:

- (i) Exactitud.
- (ii) Precisión.
- (iii) Sensibilidad analítica.
- (iv) Especificidad analítica para incluir sustancias interferentes.
- (v) Rango reportable de los resultados de examen del sistema de examen.
- (vi) Intervalos de referencia (valores normales).
- (vii) Cualquier otra característica de desempeño requerida para el desempeño del examen.

(3) Basados en las especificaciones del desempeño verificadas o establecidas de acuerdo con los párrafos (b)(1) o (b)(2) de esta sección, determinar la calibración del sistema de examen y los procedimientos de control para los exámenes de pacientes como se requiere en §493.1255 y §493.1256.

Esto significaría llevar a cabo siete experimentos diferentes:

- El experimento de comparación de métodos para estimar la inexactitud/ sesgo;
- El experimento de replicación para estimar la imprecisión;
- El experimento de linealidad para determinar el rango reportable;
- El experimento de límite de detección para estimar la concentración más baja que puede ser medida;
- El experimento de interferencia para determinar las interferencias constantes;
- El experimento de recuperación para determinar las interferencias proporcionales;
- El estudio extensivo de valores de referencia para estimar intervalo de referencia.

Estos estudios experimentales deben ser documentados. La sección §493.1253(c) requiere que el laboratorio mantenga la documentación de validación de métodos por al menos tanto tiempo como sea usado el método.

Es importante notar que estos requisitos NO son retroactivos, lo que significa que cualquier método utilizado antes de la implementación de la Regla Final de CLIA el 24 de abril de 2003 no necesita volver a ser verificado.

Reglas Específicas para la Calibración y la Verificación de la Calibración

Calibración. Existen requisitos adicionales para la calibración, como sigue:

§493.1255(a) Realizar y documentar procedimientos de calibración-

(1) Siguiendo las instrucciones del sistema de examen del fabricante, usando materiales de calibración provistos o especificados, y respetando al menos la frecuencia recomendada por el fabricante;

(2) Usando criterios verificados o establecidos por el laboratorio como se especifica en 493.1253 (b)(3) –

- (i) Usando los materiales de calibración apropiados para el sistema de examen y, si es posible, con trazabilidad a un método de referencia o a materiales de referencia de valor conocido;
 - (ii) Incluyendo el nombre, tipo y concentración de los materiales de calibración, así como los límites aceptables y la frecuencia de calibración; y
- (3) Cada vez que falle la verificación de la calibración en alcanzar los límites aceptables del laboratorio para la verificación de la calibración.

En pocas palabras, esto significa que los laboratorios deben seguir las instrucciones del fabricante para la calibración y re-calibrar tan frecuentemente como lo recomiende el fabricante, más re-calibrar cada vez que falle la verificación de calibración.

Verificación de la calibración. Este es el proceso de verificar que la calibración permanezca estable. En muchos aspectos, el procedimiento para hacer esto es similar al experimento para el rango reportable. Los requisitos específicos son los siguientes:

§493.1255(b) Realizar y documentar los procedimientos de verificación de la calibración-

- (1) Siguiendo las instrucciones de verificación de la calibración del fabricante;
- (2) Usando criterios verificados o establecidos por el laboratorio bajo 493.1253(b)(3) –
 - (i) Incluyendo el número, tipo y concentración de los materiales, así como los límites aceptables para la verificación de la calibración; e
 - (ii) Incluyendo al menos un valor mínimo (o cero), un valor medio y un valor máximo cercano al límite superior del rango para verificar el rango reportable del laboratorio de los resultados de examen para el sistema de examen; y
- (3) Al menos una vez cada 6 meses y cada vez que ocurra alguna de las siguientes situaciones:
 - (i) Es introducido un cambio completo de reactivos para un procedimiento, a menos que el laboratorio pueda demostrar que cambiar el número de lote del reactivo no afecta el rango usado para reportar los resultados de los exámenes de los pacientes, y que los valores de los controles no son afectados adversamente por cambios en el número de lote del reactivo.
 - (ii) Hay un mantenimiento preventivo mayor o un reemplazo de las partes críticas que podría influenciar el desempeño del examen.

(iii) Los materiales de control reflejan una tendencia o un cambio inusual, o están fuera de los límites aceptables del laboratorio, y otros medios de evaluar y corregir valores de los controles inaceptables fallan al identificar y corregir el problema.

(iv) La agenda establecida del laboratorio para la verificación del rango reportable para los resultados del examen del paciente requiere una verificación de la calibración más frecuente.

En resumen, los laboratorios deben ya sea verificar la calibración, o la re-calibración, al menos cada 6 meses, después de un cambio en los reactivos, después de un mantenimiento preventivo mayor, después del cambio de cualquier componente mayor, y en cualquier momento que haya un problema con los resultados del control de la calidad que persista por algún tiempo.

Requisitos Adicionales para la Validación Periódica

Finalmente, hay otros dos requisitos que podrían llevar a estudios adicionales de validación de métodos para ciertos exámenes.

- Se requiere de una evaluación en curso (al menos semi-anual) de la exactitud para cualquier examen no evaluado por un esquema de evaluación de la competencia (PT, por sus siglas en inglés) o programa de evaluación externa de la calidad.

§493.1236 Evaluación del desempeño de los esquemas de evaluación de la competencia. (c) Al menos dos veces al año, el laboratorio debe verificar la exactitud de lo siguiente: (1) Cualquier examen o procedimiento que realice no incluido en la subparte I de esta parte que es o evaluada o calificada por esquema de evaluación de la competencia aprobado por CMS. (2) Cualquier examen o procedimiento listado en la subparte I de esta parte para el que no se ofrecen muestras de esquemas de evaluación de la competencia compatibles con un esquema de evaluación de la competencia aprobado por los CMS.

- Los exámenes que son realizados por sistemas analíticos múltiples requieren comparaciones periódicas, como sigue:

§493.1281 Comparación de resultados de examen. (a) Si un laboratorio realiza el mismo examen usando diferentes metodologías o instrumentos, o realiza el mismo examen en múltiples sitios de examen, el laboratorio debe tener un sistema que, dos veces al año, evalúe y defina la relación entre exámenes utilizando diferentes metodologías, instrumentos o sitios de examen.

Estándares de la JC para la validación de métodos [3]

Exámenes Waived. Antes de 2007, los estándares de la JC incluyeron requisitos para exámenes *waived*, pero no imponen ya esos requisitos. Los laboratorios pueden sólo aceptar los reportes de desempeño del fabricante.

Exámenes Non-Waived (Complejidad Moderada y Alta) Aprobados por la FDA. Si bien JC aplica los mismos requisitos generales de control de la calidad que CLIA, la JC tiene requisitos específicos de validación de métodos (CC.1.70). La JC identifica los elementos de desempeño como sigue:

- Antes de que sea utilizado un nuevo examen, método o instrumento para reportar resultados de pacientes, el laboratorio valida que el examen, método o instrumento producirá consistentemente resultados exactos.
- El laboratorio valida la exactitud, la precisión y el rango reportable para un examen, método o instrumento no modificado aprobado para el cual el fabricante ha establecido las especificaciones de desempeño.
- El laboratorio establece especificaciones de desempeño que incluyen exactitud, precisión, sensibilidad analítica, especificidad analítica, el rango reportable, y los intervalos de referencia (valores normales) para exámenes, métodos o instrumentos modificados.

Nota: Los exámenes modificados, métodos o instrumentos incluyen lo siguiente:

- Los procedimientos de examen con modificaciones al uso aprobado por la FDA para el tipo de espécimen, los reactivos, el instrumento, los pasos del procedimiento u otros componentes
- Los exámenes o los métodos desarrollados en el laboratorio sin evaluación de la FDA
- Los exámenes, los métodos o los instrumentos no sujetos a autorización de la FDA
- Cuando un nuevo examen, método o instrumento reemplace un examen, método o instrumento anterior, se realizan las correlaciones entre el examen, el método o el instrumento anterior y el nuevo.

Nota: Si el reemplazo del examen, el método o el instrumento es debido a un desempeño pobre, se sigue el EP 1 de este estándar.

- Para cada examen nuevo, método o instrumento, el laboratorio ha determinado que los intervalos de referencia son aplicables (valores normales) para la población específica de pacientes examinada.

- El laboratorio ha documentado la validación de nuevos exámenes, métodos o instrumentos implementados.
- El laboratorio establece el número, tipo y frecuencia de los materiales de control de la calidad utilizando las especificaciones de desempeño verificadas o establecidas por el laboratorio.
- A lo largo del tiempo el laboratorio monitorea la exactitud y la precisión del desempeño del examen que podrían ser influenciadas por cambios en el desempeño del sistema de examen y en las condiciones ambientales, y variación en el desempeño del operador.

La JC acepta datos del control estadístico de la calidad e historiales de desempeño del examen como la “evidencia” de un adecuado desempeño del método cuando un método usado en una locación se instituye en otra locación.

Estándares del Programa de Acreditación de Laboratorio del CAP para validación de métodos [4]

Exámenes Waived. Como la JC, el CAP no requiere ya la validación de métodos para exámenes *waived* (ver su 9/07 LAB GEN Checklist, página 58). Sin embargo, un intervalo de referencia debe ser identificado.

Exámenes Non-Waived. La filosofía del CAP es que todos los exámenes de los laboratorios clínicos necesitan cumplir los requisitos definidos bajo la sección §493.1253(b)(2) de CLIA '88.

La lista de verificación GEN de 2007 del CAP (Laboratorio General) incluye requisitos de especificación de desempeño — exactitud y precisión (GEN:42020), rango reportable (GEN:42085), sensibilidad o límite de detección inferior (GEN:42025), e intervalo de referencia (GEN:42162) — para cada procedimiento de examen. La especificidad (GEN:42030) implica una evaluación de la habilidad del método para responder correctamente a la concentración del analito en la presencia de sustancias interferentes. La lista de verificación de CAP ahora establece:

“¿Ha verificado o establecido y documentado el laboratorio las interferencias analíticas para cada examen? Las sustancias interferentes plantean un problema significativo al laboratorio clínico y a los profesionales de la salud que podrían ser desorientados por los resultados del laboratorio que no reflejan el estado clínico del paciente. El laboratorio debe considerar las interferencias comunes realizando estudios o teniendo a su disposición estudios realizados en algún otro lado (tales como por el fabricante para el reactivo del instrumento)”.

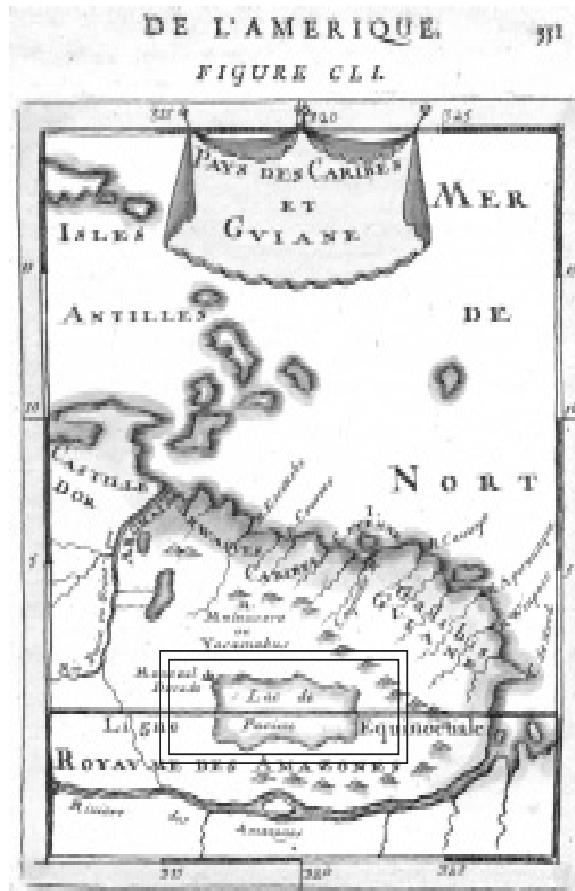
¿Cuál es el punto?

Las regulaciones de la Regla Final de CLIA establecen estándares *mínimos* para la gestión de la calidad de los procesos de examen. Otras agencias de inspección y organizaciones deben proveer estándares que sean *al menos tan demandantes como los de CLIA* para ser aprobados por CMS para la inspección y la acreditación de laboratorios de EUA. Desde 2008, estas acreditadoras han adaptado sus estándares para ser consistentes con CLIA, en algunos casos, *disminuyendo o eliminando* sus requisitos anteriores.

- Para métodos *non-waived* aprobados por la FDA, usted debe validar la precisión, la exactitud, el rango reportable, y los intervalos de referencia para cualquier nuevo examen o sistema analítico.
- Para métodos *non-waived* NO aprobados por la FDA o modificados por el laboratorio, debe validar las características adicionales de sensibilidad analítica y especificidad analítica.
- Además, cada seis meses debe validar la exactitud de los métodos donde no esté disponible un esquema de evaluación de la competencia aprobado, comparar los resultados de examen si hay múltiples métodos en servicio, y verificar la calibración cada vez que hay cambios significativos en reactivos, componentes y desempeño (problemas evidenciados en los datos del control estadístico interno de la calidad).

Referencias

1. U.S. Department of Health and Human Services. Medicare, Medicaid and CLIA programs: Regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. Fed Regist 1992; 57:7002-186.
2. US Centers for Medicare & Medicaid Services (CMS). Medicare, Medicaid, and CLIA Programs: Laboratory Requirements Relating to quality Systems and Certain Personnel Qualifications. Final Rule. Fed Regist Jan 24 2003;16:3640-3714. See also the CMS website <http://www.cms.hhs.gov/clia/> and the CMS State Operations Manual Appendix C, Regulations and Interpretive Guidelines for Laboratories and Laboratory Services, http://www.cms.hhs.gov/CLIA/03_Interpretive_Guidelines_for_Laboratories.asp
3. Accreditation Manual for Pathology and Clinical laboratory Services.Joint Commission (JC). Oakbrook Terrace, IL. 60181
4. CAP Laboratory Accreditation Checklists. College of American Pathologists (CAP). 325 Waukegan Road, Northfield, IL 60093-2750 <http://www.cap.org>
5. COLA Accreditation Manual. 9881 Broken Land Parkway, Suite 200, Columbia, MD 21046. <http://www.cola.org>



La ciudad mítica del oro: El Dorado

PAYS des CARIBES et GVIANE, by Alain Mallet, Paris 1683. The mythical city of gold – Manoeal Dorado – on the western bank of the mythical Lac De Parime.

Se rumoreaba ampliamente que el Lago de Parima y El Dorado (ciudad del oro) existían y esto estaba bien documentado en los mapas de esa época, pero nunca fueron implementados (descubiertos)- algo así como las regulaciones de CLIA en E.U.A. que describen un proceso de compensación del control de la calidad cuya implementación aún se mantiene demorada. La Regla Final de CLIA (24 de Enero del 2003) puso bien en claro que el proceso de compensación del control de la calidad es solo un mito. La gente buscó, espero ansiosamente por la nueva implementación y buscó algo más, pero la ciudad del oro nunca apareció.

5: ¿Cómo es seleccionado un método?

El proceso de establecer un examen de rutina comienza con la selección del método a ser evaluado o validado. El Dr. Westgard describe diferentes tipos de características de un método e identifica aquellas importantes en la selección y validación de métodos. Ejemplos de estas características están incluidos para un examen de colesterol.

Objetivos:

- Definir los pasos para establecer un examen de rutina de laboratorio.
- Identificar los diferentes tipos de características de un método.
- Reconocer las características que son el enfoque de los estudios de validación de métodos.

Materiales del capítulo:

- VM — **Seleccionando un método para validar,**
por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar la lección.
- Relacionar los tipos de características de un problema común de selección, tales como encontrar un lugar donde alojarse, comprar un nuevo abrigo, decidir a dónde ir para cenar, etc.

Preguntas de Autoevaluación:

- ¿Cuáles son los tres tipos de características de un método?
- De tres ejemplos de cada uno de los tipos de características de un método.
- ¿Cuáles son las características de mayor interés al momento de seleccionar un método?
- ¿Cuáles son las características de mayor interés durante la validación de un método?

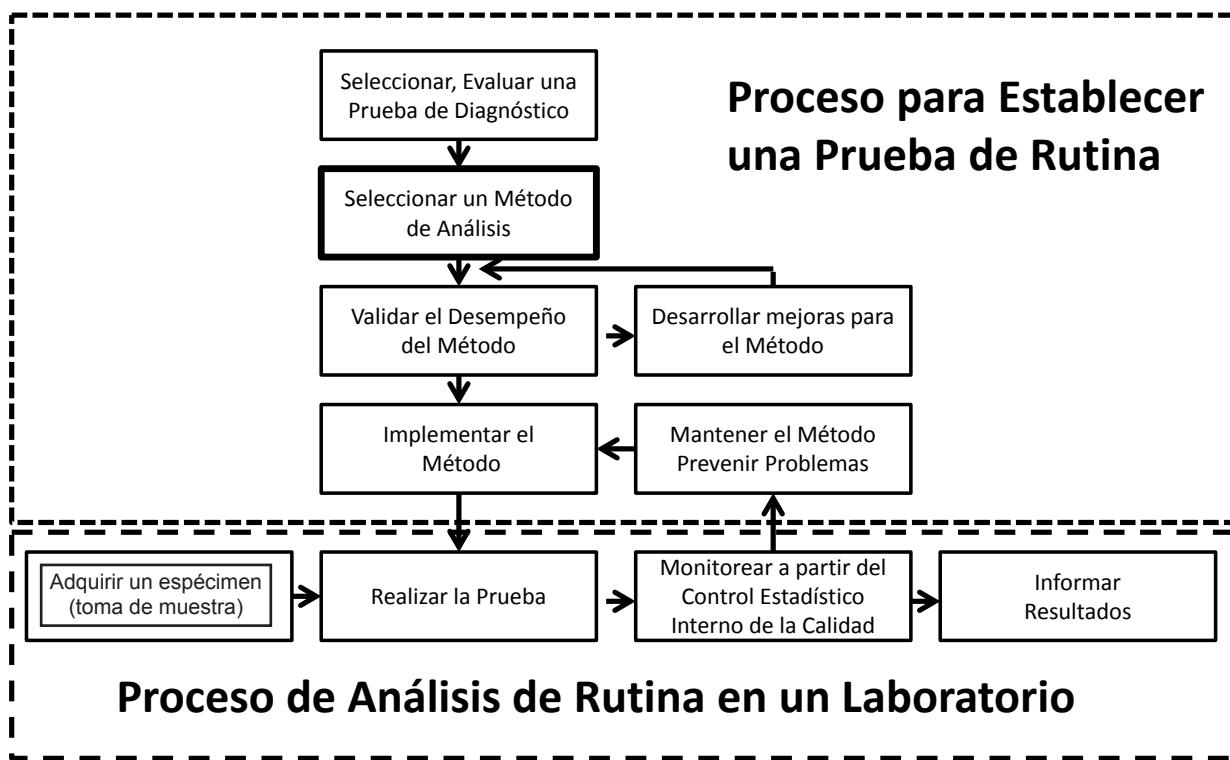
Validación de Métodos: Seleccionando un Método para Validar

James O. Westgard, PhD

La validación de métodos se trata de evaluación de errores, como se discutió antes en *El secreto más interno oculto y profundo sobre la validación de métodos*. Sin embargo, antes de que pueda evaluar cualquier error, tiene que *seleccionar* el método a ser validado. La *selección* de métodos es un proceso diferente que necesita ser entendido en relación al proceso de validación que seguirá. De hecho, hay varios procesos que son esenciales para establecer un método de análisis de rutina.

Estableciendo un proceso de examen de laboratorio

En la figura de abajo se muestran las actividades importantes para establecer un método de rutina de análisis. Los cuadros mostrados en la porción inferior del gráfico ilustran los pasos clave involucrados en el análisis de rutina, donde el laboratorio adquiere especímenes, realiza exámenes, chequea el control estadístico interno de la calidad, y reporta los resultados del laboratorio. Esas actividades son generalmente referidas como el trabajo real del laboratorio.



Sin embargo, para que un análisis se convierta en rutina, las otras actividades mostradas en la figura son muy importantes. La selección de los exámenes de diagnóstico es de hecho el primer paso, pero esto es a menudo omitido para los exámenes comunes cuya utilidad clínica es bien aceptada. Para estos exámenes establecidos, usualmente comenzamos con la selección del método (el cuadro con el borde más oscuro), luego validamos su desempeño. Si el desempeño es aceptable, el método es implementado para el servicio de rutina. Si el desempeño no es aceptable, el laboratorio podría desarrollar algunas mejoras, aunque eso se convierte en algo cada vez más difícil con el alto grado de automatización de muchos sistemas analíticos. Hoy en día es más probable que un laboratorio seleccione otro método en lugar de intentar hacerle mejoras, luego comenzará el proceso de validación una vez más para el nuevo método.

Una vez que un método ha demostrado un desempeño aceptable, el método debe ser implementado para operaciones de rutina. Esto involucra definir los procedimientos estándar de operación y documentar el procedimiento, seleccionando un procedimiento de Control Estadístico Interno de la Calidad apropiado para el monitoreo del desempeño de rutina, y entrenar al personal para operar el nuevo método. En un servicio de rutina, los problemas serán sin duda identificados a través del Control Estadístico Interno de la Calidad, lo que llevará a procedimientos de mantenimiento preventivos para minimizar o eliminar esos problemas. La operación de rutina es a menudo la parte más simple de todo este proceso si el laboratorio hace un buen trabajo al seleccionar el método, valida el desempeño del método, implementa el método a través de entrenamiento en servicio cuidadoso y minucioso, monitorea el desempeño del método con un procedimiento de Control Estadístico Interno de la Calidad que tiene una tasa baja de falso rechazo y una detección apropiada de errores, y mantiene activamente el método para identificar problemas, eliminar fuentes de error, y prevenir futuros problemas.

Características del método

La meta cuando se selecciona un método es escoger el método que tiene la mejor oportunidad de alcanzar los requisitos de servicio del laboratorio. El proceso de selección consiste en definir aquellos requisitos, buscando la literatura técnica para encontrar información acerca de métodos disponibles, luego seleccionar el método cuyas características satisfacen mejor los requisitos de servicio del laboratorio.

Una definición cuidadosa de los requisitos es esencial [1]. Si se omite este paso, el laboratorio podría gastar un tiempo y esfuerzo considerables evaluando un método que no será satisfactorio, sin importar el desempeño del método. Un ejemplo podría ser un método de *point-of-care* que resulta ser muy costoso aunque su desempeño analítico sea aceptable. El costo debería ser considerado

durante el proceso de selección del método a ser evaluado, no después de validar el desempeño del método.

¿Qué características de un método son importantes? En general, pueden ser divididas en tres categorías.

- **Las características de aplicación** son factores que determinan si un método puede ser implementado en una situación particular del laboratorio. Consisten en un costo-por-examen, tipos de especímenes que pueden ser analizados, volumen de la muestra, tiempo de entrega de resultados, carga de trabajo, requisitos de equipo y personal, espacio, transporte y consideraciones de seguridad.
- **Las características metodológicas** son factores que, en principio, deberían contribuir a un mejor desempeño. En general, tienen que ver con la sensibilidad analítica y la especificidad analítica del método de análisis. Consideran la elección de la reacción química, la optimización de las condiciones de reacción, los principios de estandarización y calibración, y el rigor del procedimiento analítico.
- **Las características de desempeño** son factores que, en la práctica, demuestran qué tan bien se desempeña un método. Incluyen el rango reportable, la precisión, la recuperación, la interferencia, la exactitud y algunas veces el límite de detección.

Comencemos con un ejemplo fuera del laboratorio que ilustra estos diferentes tipos de características. Asuma que está seleccionando un vehículo motorizado nuevo.

- **Las características de aplicación** incluyen su uso previsto, el tipo de vehículo, el rango de precios, el estilo del cuerpo, el color, etc. Su uso previsto ayudará a definir el tipo de vehículo—tractor, camión, automóvil o motocicleta. Asumiendo que es un vehículo que se usará para ir al trabajo, busquemos un automóvil. Si su color favorito es el rojo, no tiene que probar con un coche negro—puede saber por la vista si satisface esta característica de aplicación.
- **Las características metodológicas** pueden ser ilustradas por lo siguiente. Si usted vive en Wisconsin, o peor aún en Minnesota (sin ofender; he vivido ahí y aún visito a menudo), podría querer un automóvil con buena tracción por el clima en invierno. Se esperaría que un vehículo con tracción en las cuatro ruedas funcione mejor que uno con tracción en dos ruedas. Si también fuera importante obtener un buen rendimiento por kilometraje, un motor de cuatro cilindros sería una mejor opción que un motor de seis u ocho cilindros. Tanto la elección de un vehículo con

tracción en la cuatro ruedas como el número de cilindros (o tamaño del motor) son características metodológicas que deberían ayudar a satisfacer sus necesidades de desempeño.

- **Las características de desempeño** son aquellos factores que puede evaluar al conducir el vehículo. Las medidas estándar de desempeño son la aceleración, el manejo, el confort de conducción, los frenos y el kilometraje. Las mediciones especiales podrían incluir tracción fuera del camino o cambio de tracción para clima invernal. Usted no hace una prueba de manejo en un camión costoso, grande, negro, de ocho cilindros y tracción en dos ruedas porque el vehículo no satisface su aplicación y sus características metodológicas.

Ejemplo de colesterol

Las características generales por considerar son las siguientes:

- **Las características de aplicación** deben incluir el tipo de espécimen, el volumen del espécimen, la carga de trabajo apropiada para la situación de examen (un laboratorio centralizado de alto volumen contra un *point-of-care* de bajo volumen), el costo por examen, el tiempo para el análisis, las habilidades del operador, y los requisitos de entrenamiento del operador.
- **Las características metodológicas** deberían incluir el tipo de estándares o calibradores, la trazabilidad del estándar o de los valores asignados del calibrador, el principio químico, los reactivos y las condiciones de reacción, el principio de medición, y las capacidades de medición.
- **Las características de desempeño** deberían incluir el rango reportable, la imprecisión, el sesgo contra el método de referencia Abell-Kendahl [3], la interferencia, y un error total menor al 10%.

Las características de un laboratorio automatizado de alto volumen pueden ser definidas más específicamente:

- **Las características de aplicación** son muestras de suero, un tamaño de muestra menor a 10 microlitros, un rendimiento de examen de 200 pruebas por hora, un costo menor que \$0.25 por “click” o \$0.50 por examen.
- **Las características metodológicas** son la reacción enzimática de esterasa de colesterol, calibradores con trazabilidad al método de referencia nacional Abell-Kendahl [3], y una exactitud verificada por comparación de resultados con un laboratorio de referencia en lípidos.

- **Las características de desempeño** son un rango reportable de 50 a 500 mg/dL, un error total permitido del 10% en una concentración de 200 mg/dL, un CV para el método de 3% o menor, un sesgo para el método del 3% o menos, sin interferencia por ictericia hasta una concentración de bilirrubina de 10 mg/dL, sin interferencia por lipemia hasta una concentración de trigliceridos de 600 mg/dL, y una interferencia mínima por hemólisis.

Las características de un *point-of-care* (POC) podrían ser bastante diferentes, particularmente las características de aplicación que deben tomar en cuenta los sitios POC:

- **Las características de aplicación** son un espécimen de sangre entera obtenida por punción capilar o una muestra heparinizada, un tiempo de entrega de 10 minutos o menos, facilidad de uso para operadores que no están entrenados en exámenes de laboratorio, un mantenimiento y tiempo muerto mínimo, funcionen o no funcionen, un costo menor a \$4.00 por examen, un lote de reactivos imposible de usar no calibrado, pequeño, ligero y portátil.
- **Las características metodológicas** son la reacción enzimática de esterasa de colesterol, una calibración incorporada en el sistema de examen por el fabricante con valores específicos de lote, todos con trazabilidad al método de referencia Abell-Kendahl [3], y una exactitud verificada por comparación de resultados con un laboratorio de referencia en lípidos.
- **Las características de desempeño** son un rango reportable de 100 a 400 mg/dL, un error total permitido del 10 % en una concentración de 200 mg/dL, sin interferencia con hematocritos de hasta 55 %, sin interferencia por bilirrubina de hasta 5 mg/dL, sin interferencia por una lipemia de hasta 600 mg/dL de triglicéridos, y una interferencia mínima por hemólisis.

Selección de sistemas analíticos de exámenes múltiples

Se torna mucho más complicado cuando se intenta seleccionar un sistema analítico que lleva a cabo muchos exámenes diferentes. Los sistemas analíticos hoy en día pueden ofrecer un centenar de exámenes diferentes. Son estos instrumentos enormes, que pueden consolidar mucho del trabajo de laboratorio de una sola “caja”. Cuando el objetivo es la consolidación, entonces las características de aplicación dominarán el proceso de selección. Las características metodológicas podrían ser de interés para algunos exámenes críticos, pero la calidad analítica de muchos exámenes será *asumida* como aceptable. A pesar de esto, el desempeño de cada examen debe ser validado para rango reportable, precisión, exactitud e intervalos de referencia.

Como un ejemplo de las características de aplicación en el “mundo real”, una presentación de un trabajo libre[4] en una convención nacional se enfocó en las siguientes características: el número de diferentes exámenes disponibles (menú de pruebas), los exámenes/hora (rendimiento), cantidad de reactivos a bordo, la carga de reactivos, la capacidad de carga de muestras, el volumen de la muestra, el volumen muerto mínimo, la detección de coágulos/burbujas, el tiempo de análisis, la demora durante la puesta en marcha, el tiempo hasta generar el primer resultado, la capacidad para muestras de urgencia y el tiempo de demora para las urgencias, las horas de mantenimiento diario, el registro de mantenimiento realizado en el software del equipo, la simplicidad de operación del software, las reglas de Control de la Calidad disponibles y el tiempo de entrenamiento. Éstas son características de aplicación complejas que requieren un estudio cuidadoso y minucioso para comparar las características de diferentes sistemas.

Encontrar toda la información es un reto. Los fabricantes a menudo proveen descripciones para sus sistemas analíticos en sus sitios web y en sus materiales de mercadeo. La comparación de las características de diferentes sistemas analíticos es ayudada en gran medida por “reportes de comparación de instrumentos”, tales como los provistos en *CAP Today* del CAP “College of American Pathologists”. Existen enormes exhibiciones de equipos y sistemas en convenciones de profesionales tales como la de la AACC “American Association of Clinical Chemistry”. Finalmente, y tal vez lo más importante antes de la decisión final de compra, debe haber visitas a los laboratorios donde los sistemas ya están instalados y funcionando.

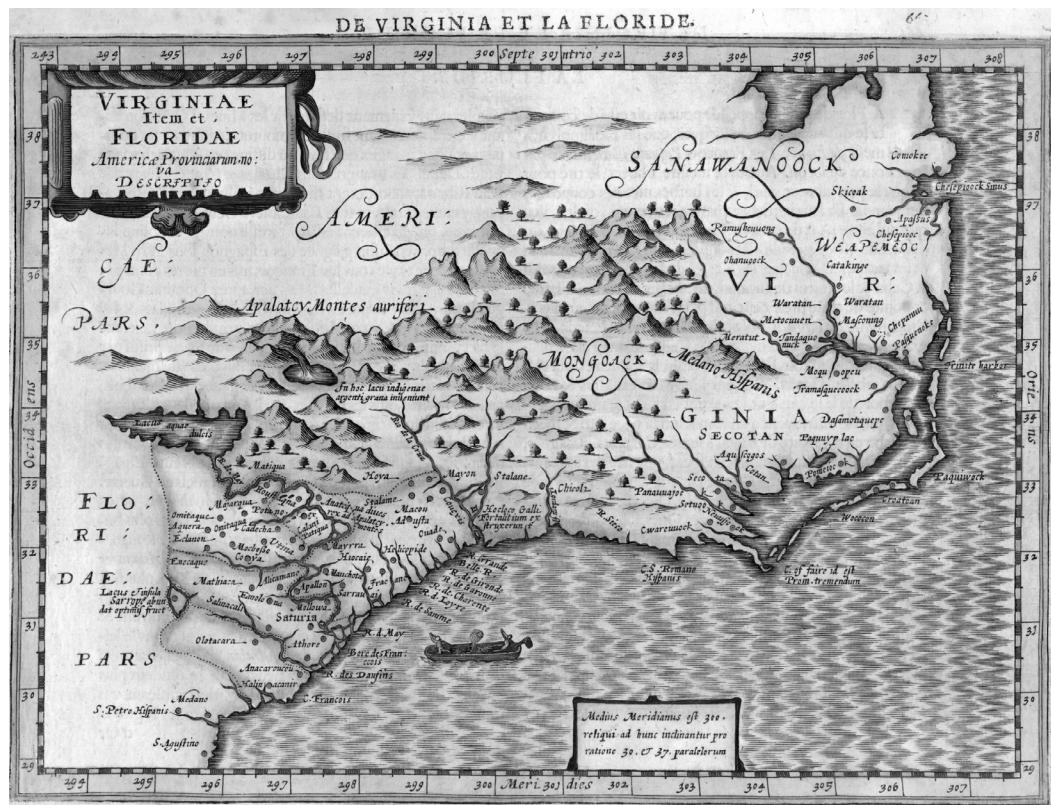
La información y las especificaciones para el desempeño del método podrían ser más difíciles de encontrar. Si bien se solicita a los fabricantes declarar sus especificaciones de desempeño para obtener aprobación de mercado de la FDA, estos reportes sobre las características de desempeño podrían ser difíciles de encontrar en su material de mercadeo. A veces los estudios de validación de métodos son publicados en la literatura científica, pero esto está usualmente limitado a nuevos exámenes y nuevos sistemas. Es más común encontrar buena información científica en los resúmenes y carteles presentados en convenciones profesionales. Para los sistemas analíticos que tienen algo de historia en el campo, su desempeño comparativo podría ser consultado en reportes de encuestas de esquemas de evaluación de la competencia y programas de evaluación externa de la calidad de organizaciones tales como el CAP, el *Medical Laboratory Evaluation* (MLE), y la *American Association of Bioanalysts* (AAB).

¿Cuál es el punto?

La selección de un método o de un sistema analítico lleva directamente a la calidad de los exámenes disponibles de un laboratorio. Por tanto, el proceso de selección en sí mismo debe ser objetivo y justificar la aplicación y las características metodológicas que son importantes en su laboratorio. El proceso de selección es particularmente complicado para los sistemas analíticos de exámenes múltiples, donde las características de aplicación dominarán la decisión. Para mantener la calidad analítica presente en el proceso de decisión, podría ser necesario ir más allá de la consideración de las características metodológicas y observar el desempeño comparativo de los métodos, información disponible en los reportes de encuestas de los esquemas de evaluación de la competencia y programas de evaluación externa de la calidad. El laboratorio debe validar las características de desempeño de todos los exámenes en el sistema. Y el laboratorio poseerá la calidad que comporte.

Referencias

1. Westgard JO, DeVos DJ, Hunt MR, Quam EF, Carey RN, Garber CC. Concepts and practices in the evaluation of laboratory methods. I. Background and Approach. *Am J Med Technol* 1978;44:290-300.
2. Wiebe DA, Westgard JO. Cholesterol - a model system to relate medical needs with analytical performance. *Clin Chem* 1993;39:1504-1513.
3. Cooper GR, Smith SJ, Duncan IW, et al. Interlaboratory testing of the transferability of a candidate reference method for total cholesterol in serum. *Clin Chem* 1986;32:921-929.
4. Perkins SL, Bookalam S, Warr M, Ooi DS. Evaluation of the Architect i1000SR Immunoassay Analyzer. Abstract D-81. *Clin Chem* 2007/53:A185.



Un problema de selección!

**VIRGINIAE ITEM et FLORIDAE, by Johannes Cloppenburg,
Amsterdam, 1630.**

Este mapa de la costa sudeste de E.U.A. fue preparado a partir de dos mapas anteriores – uno de Florida y uno de Virginia. Si bien el mapa fue bien realizado, es decir, un grabado hermoso, no fue de mucha utilidad porque omitió una característica muy importante – Las Carolinas. En consecuencia las islas “Cape” del lado oeste del mapa de Virginia se unieron a la geografía del este de Florida, sin dejar lugar para Las Carolinas. Un mito notorio surgido de este mapa es el lago no existente en las montañas *Appalachian*, que fue conocido como el Lago *Appalachia*.

6: ¿Qué experimentos son necesarios para validar el desempeño de un método?

El Dr. Westgard explica cómo desarrollar un plan experimental para un estudio de validación de métodos. Pueden ser utilizados experimentos específicos para estimar la magnitud de los diferentes tipos de errores analíticos. Estos experimentos pueden ser organizados en un plan práctico en base al tiempo y esfuerzo necesarios para llevar a cabo los diferentes experimentos.

Objetivos:

- Identificar los diferentes experimentos utilizados en un estudio de validación de métodos.
- Relacionar esos experimentos a los tipos de error estimados.
- Organizar esos experimentos en un plan práctico para llevar a cabo el estudio.

Materiales del capítulo:

- VM — El plan experimental, por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar la lección.
- Revisar un reporte de literatura y sus procedimientos experimentales.

Preguntas de Autoevaluación:

- ¿Qué experimento es utilizado para estimar la imprecisión de un método?
- ¿Qué experimentos son utilizados para estimar la inexactitud de un método?
- ¿Por qué el experimento para rango reportable es realizado al comienzo del estudio de validación?
- ¿Por qué por lo general se realizan dos experimentos de replicación?
- ¿Qué diferencia existe entre el error sistemático proporcional y constante?
- ¿Qué experimentos se realizan para estimar los errores sistemáticos proporcional y constante?

Validación de Métodos: El Plan Experimental

James O. Westgard, PhD

Comienzo esta disertación asumiendo que el método a ser evaluado ha sido seleccionado cuidadosamente, como se discutió anteriormente en *Seleccionando un método para validar*. Por lo tanto, los requisitos de aplicación han sido satisfechos y las características metodológicas han sido consideradas.

Ahora podemos enfocarnos en las características de desempeño, los cuales incluyen la precisión, la exactitud, la interferencia, el rango reportable, y algunas veces los límites de detección. Estas características podrían haber sido ya estimadas por el fabricante (especificaciones de desempeño establecidas para el método) o por un usuario (publicados en un estudio de validación). Estos reportes o resultados publicados aún necesitan ser verificados para demostrar que el método funciona adecuadamente y que es aceptable en su laboratorio particular. Ese es el propósito de un estudio de validación de métodos.

Acercamiento para la formulación de un plan

Para llevar a cabo un buen estudio de validación de métodos, necesita hacer lo siguiente:

- definir un requisito de la calidad para el examen en la forma de la cantidad de error que es permitida, preferentemente un error total permitido;
- seleccionar experimentos apropiados para revelar los tipos de errores analíticos esperados;
- recolectar los datos experimentales necesarios;
- realizar cálculos estadísticos a partir de los datos para estimar el tamaño de los errores analíticos;
- comparar los errores observados con el error permitido definido, y
- juzgar la aceptabilidad del desempeño del método observado.

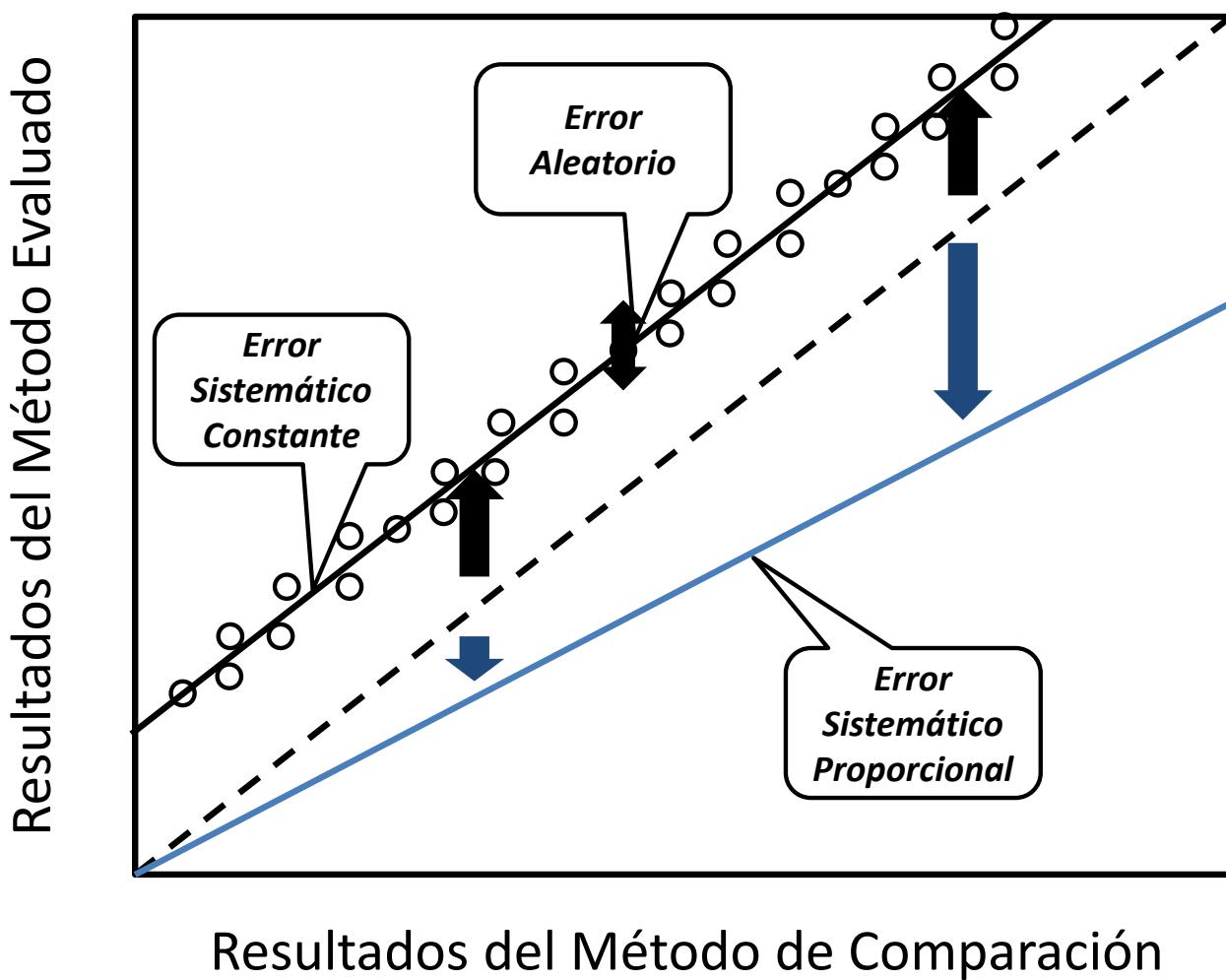
Un plan experimental puede ser formulado al:

- reconocer los tipos de errores que necesitan ser evaluados para este examen y este método;
- identificar los experimentos apropiados y la cantidad de datos necesaria para estimar esos tipos de error, luego
- organizar estos experimentos para realizar los que son rápidos y sencillos primero, y los que toman más tiempo y esfuerzo después.

Tipos de errores que deben ser evaluados

Todas las mediciones tienen algún error! Incluso los instrumentos simples de medición, tales como una báscula de baño, tienen errores. Cada vez que se pesa en una báscula de baño, observa error. Por eso es que usted se baja inmediatamente de la báscula, se sube nuevamente y hace otra medición. Usualmente observa que estas mediciones, aunque realizadas próximas en el tiempo y esencialmente bajo condiciones idénticas, no son aun exactamente las mismas — esto es provocado por el error aleatorio o imprecisión como hemos discutido anteriormente en *El secreto más interno oculto y profundo sobre la validación de métodos*. Podría también haber notado que virtualmente todas las escalas son inexactas —todas marcan muy alto, ¿no es cierto? Este es un ejemplo del error sistemático o sesgo descrito anteriormente.

En respuesta a que consideramos que el peso obtenido es muy alto, usualmente tratamos de ajustar el punto cero de la báscula y lograr los resultados más bajos. Esto asume que hay un error sistemático que es constante en la



naturaleza, esto es, toda la gente que se pese en esa báscula tendrá mayor peso en la misma cantidad. Si en lugar de ello los pesos de todas las personas tienen error en proporción a su peso total, esto es, error proporcional, la medición necesita ser corregida mediante un tipo de calibración de ajuste, más que un ajuste de punto cero.

Si artefactos tan simples como las básculas, están sujetos a error, es totalmente comprensible que las mediciones de los dispositivos complejos y sistemas utilizados para los exámenes de laboratorio estén sujetos a los mismos tipos de error:

- imprecisión o errores aleatorios,
- inexactitud, sesgo, o errores sistemáticos, que pueden ser de dos tipos
 - error sistemático constante, o
 - error sistemático proporcional.

Todos estos errores pueden ser reconocidos cuando un grupo de mediciones es comparado con los valores correctos o verdaderos. Por ejemplo, la figura de arriba muestra cómo los diferentes tipos de error se revelan cuando los resultados de un método de examen son graficados en el eje "y" contra aquellos obtenidos a partir de un método de comparación en el eje "x". La línea punteada en el medio de la figura representa el desempeño del método ideal donde el método de examen y el método de comparación dan exactamente los mismos resultados. La línea inferior en la figura muestra el efecto de un error sistemático proporcional, donde la magnitud del error se ve incrementada en cuanto el resultado del examen se torna más alto. La línea superior muestra el efecto de un error sistemático constante, donde toda la línea se desplaza hacia arriba y todos los resultados son mayores en la misma cantidad. Nótese que estos resultados también están afectados por el error aleatorio del método, por consiguiente, los puntos de los datos reales estarían dispersos alrededor de la línea ilustrada en la figura. El rango generado por la dispersión hacia arriba y abajo de la línea brinda una idea de la cantidad de error aleatorio que esté presente.

Experimentos para estimar los errores analíticos

Si bien un experimento de comparación de métodos puede revelar todos estos diferentes tipos de errores, no es necesariamente la mejor manera de ir y estudiar un nuevo método. Por ejemplo, el error aleatorio podría ser estimado más rápidamente por análisis de un solo espécimen o un material de control estable.

Hay experimentos específicos para la estimación de los diferentes tipos de error analítico, como se muestra en la tabla abajo. La primera columna enumera el tipo de error. La segunda columna está rotulada como "preliminares" porque

Tipo de Error Analítico	Experimentos de Evaluación	
	Preliminares	Finales
Error Aleatorio	Replicación Intra corrida	Replicación Entre corridas
Error Sistemático Constante	Interferencia	Comparación de Métodos
Error Sistemático Proporcional	Recuperación	

estos experimentos son generalmente más sencillos de realizar y toman menos tiempo y esfuerzo que los experimentos “finales”. Los experimentos finales son más demandantes y deben ser realizados después de que los resultados de los experimentos preliminares han mostrado que todo es aceptable hasta ahora. Sin embargo, si se observa un desempeño pobre a partir de los datos obtenidos en los experimentos preliminares se puede detener el estudio y rechazar un método debido a que una condición de error específica ha sido identificada.

Aquí hay una pequeña descripción de estos diferentes experimentos:

- **Un experimento de replicación** provee información acerca del error aleatorio y se lleva a cabo haciendo mediciones de una serie de alícuotas de las mismas muestras dentro de un periodo de tiempo específico, usualmente dentro de una corrida analítica, en un día o durante un periodo de un mes. El experimento preliminar usualmente involucra la determinación de la imprecisión intracorrida (condiciones de repetibilidad). El experimento final generalmente requiere de al menos 20 días de trabajo para generar un buen estimado de la imprecisión total (precisión intermedia o intralaboratorio), que incluye los componentes de precisión intra e intercorridas y entre días.
- **Un experimento de interferencia** provee información acerca del error sistemático constante causado por la falta de especificidad del método. Se prepara una muestra para examen adicionando el material sospechoso a una muestra que contenga el analito. Una segunda alícuota de la muestra original es diluida con la misma cantidad de un diluyente, luego ambas muestras son analizadas por el método de examen y se determina la diferencia.

- **Un experimento de recuperación** provee información acerca del error sistemático proporcional causado por una reacción competitiva. Se prepara una muestra para examen añadiendo una solución estándar del analito que se está midiendo a una alícuota del espécimen de un paciente. Se prepara una muestra control añadiendo una cantidad igual del solvente usado para la solución estándar a una segunda alícuota del mismo espécimen del paciente. Las dos muestras son entonces analizadas por el método de examen y la cantidad recuperada es comparada a la cantidad añadida.
- **Un experimento de comparación de métodos** es primariamente usado para estimar el error sistemático promedio observado con muestras reales de pacientes, pero también puede revelar la naturaleza constante o proporcional de ese error. Se recolecta una serie de especímenes de pacientes y se analizan tanto por el método de examen como por un método analítico de comparación. Los resultados son comparados para determinar las diferencias entre los métodos, que representan los errores analíticos entre los métodos.

Organizar los experimentos en un plan

Un plan general para validar el desempeño de un nuevo método se bosqueja abajo. Incluye cuatro fases —la familiarización inicial con el método, los experimentos preliminares de evaluación “rápidos y sucios”, seguidos de estudios más extensivos de precisión y exactitud, y concluyendo con los pasos para implementar el método para el trabajo de rutina.

Periodo de familiarización

- Establecer el procedimiento de trabajo
- Validar el rango reportable
- Verificar la calibración
- Verificar el límite de detección

Experimentos preliminares de Validación de Métodos (VM por sus siglas en inglés)

- Realizar estudios de replicación intracorrida
- Realizar estudios de interferencia
- Realizar estudios de recuperación
- Juzgar la aceptabilidad analítica

Experimentos finales de Validación de Métodos (VM por sus siglas en inglés)

- Realizar estudios de replicación total
- Realizar estudios de comparación de métodos
- Juzgar la aceptabilidad analítica
- Verificar el(os) intervalo(s) de referencia
- Documentar los estudios.

Implementación

- Seleccionar los procedimientos de Control Estadístico Interno de la Calidad
- Escribir los protocolos o procedimientos de operación
- Entrenar a los analistas
- Introducir el método en el servicio
- Monitorear el desempeño de rutina

Nótese que este plan puede y debería ser ajustado para considerar alguna característica única de un método o algún requisito especial de un laboratorio y de los pacientes a quienes sirve. Por ejemplo, los estudios de interferencia y recuperación podrían ser más extensivos si el hospital es un centro oncológico cuyos pacientes serán tratados probablemente con varios medicamentos diferentes. En un centro de trasplantes, los estudios de comparación de métodos podrían enfocarse en los pacientes trasplantados que están en tratamiento con fármacos anti rechazo. La cantidad de datos necesarios también puede ser ajustada en base a lo que ya está disponible en la literatura o a lo que se requiere por lineamientos regulatorios o de acreditación. Por ejemplo, este plan debería tomar en consideración la complejidad de los métodos, como se clasifica en los lineamientos de la Regla Final de CLIA. Cualquier método desarrollado en casa o cualquier método de fabricante modificado requieren estudios más cuidadosos y podrían aplicarse los protocolos de CLSI para el experimento de replicación y para los experimentos de comparación de métodos. Los métodos sin modificación *non-waived* podrían ser evaluados con experimentos más simples que requieren menos tiempo y esfuerzo y también menos datos. Para métodos *non-waived*, no modificados aprobados por la FDA, existe un protocolo más simple planeado para la verificación por parte del usuario para veracidad y la precisión en el laboratorio, el cual será discutido en el capítulo 19.

Un recorrido por el plan

El primer paso con cualquier método nuevo es hacer que el método funcione y establecer un protocolo de operación. Éste es el periodo de “familiarización”, donde el objetivo es aprender cómo trabajar apropiadamente el método y establecer un protocolo de operación que genere resultados de examen consistentes. Con nuestra báscula de baño, por ejemplo, tiene que sacar la báscula fuera de la caja, transportarla al lugar apropiado, encontrar un espacio para ponerla, y probarla. Con un nuevo método analítico, tiene que instalar el instrumento, preparar los reactivos, calibrar los métodos y obtener resultados a partir del examen de las muestras. Uno de los factores críticos es revisar los estándares y confirmar que el método está calibrado apropiadamente, de otro modo los errores de calibración aparecerán a lo largo del trabajo experimental.

Una vez que el método esté en operación, el siguiente paso es determinar el rango reportable. Con nuestra báscula de baño, generalmente nos preocupa que el rango sea adecuado para poder pesar desde un niño pequeño hasta un adulto moderadamente grande. La báscula común no tiene probabilidad de ser suficientemente sensible para pesar a un pequeño bebé, así que si un límite de detección muy bajo es importante, esta característica necesita ser enfrentada directamente cuando se seleccione la báscula. La báscula común tampoco tiene probabilidad de poder pesar un jugador de futbol americano de más de 130 kilos, así que si es importante un rango de trabajo extremadamente amplio, esa característica, de nuevo, debería ser considerada durante el proceso de selección. Del mismo modo, con un nuevo método analítico, el rango reportable variará de examen a examen y debe ser definido como parte de las especificaciones del método, luego verificado analizando una serie de soluciones, usualmente en duplicado o triplicado, cuyas concentraciones cubran todo el rango analítico de interés. Si el límite de detección es una característica crítica, podría ser evaluado en este momento o en la siguiente fase de experimentos preliminares.

Después se realizarían los experimentos preliminares para determinar la imprecisión intracorrida, recuperación e interferencia. Se recolectan cantidades de datos mínimas en un período corto de tiempo para facilitar una evaluación rápida sobre la aceptabilidad del desempeño del método bajo las condiciones más simples. El experimento de replicación podría incluir 20 muestras de dos o tres materiales cuyas concentraciones que estén próximas a los niveles de decisión médica de interés para los exámenes. Los experimentos de interferencia deberían examinar problemas comunes tales como la hemólisis, la lipemia y altas concentraciones de bilirrubina. Los experimentos de recuperación evalúan si existe alguna reacción competitiva debida a la matriz o a otros materiales presentes en los especímenes originales. Si los errores revelados por los experimentos preliminares son pequeños, se deben realizar los experimentos finales de replicación y comparación de métodos.

Este experimento final de replicación debería cubrir al menos 20 días de trabajo. El experimento de comparación de métodos será realizado usualmente con especímenes frescos de pacientes, pero también se deberían examinar especímenes almacenados si esas condiciones de almacenamiento representan un proceso y manejo típico de los especímenes analizados rutinariamente. El método comparativo debería, idealmente, ser uno que haya sido estudiado previamente y un mínimo de 40 muestras de pacientes bien escogidas deberían ser examinadas por un mínimo de 5 días de trabajo. Estas muestras deberían estar distribuidas en un tercio del rango bajo a normal-bajo, un tercio en el rango normal, y un tercio en el rango anormal alto. Una vez que se han recolectado estos datos, la aceptabilidad del método debería ser evaluada en base al tamaño de los errores analíticos aleatorios, sistemáticos y totales. Si estos errores son pequeños en comparación con la cantidad de error que invalidaría el uso e interpretación de un resultado de examen, el método es aceptable. Si es muy grande, estos errores harán necesario rechazar el método o identificar y eliminar las causas de los errores.

Aún cuando el desempeño de un método sea evaluado como aceptable, podría ser necesario estimar o al menos verificar el o los intervalos de referencia. Si el desempeño del método comparativo no está bien documentado, podría también ser necesario realizar estudios clínicos para correlacionar los resultados de los exámenes con las condiciones clínicas. Finalmente, todos estos estudios necesitan ser documentados para futuras referencias.

La implementación comienza escribiendo el protocolo del método o el procedimiento de laboratorio que será usado para entrenar a otros analistas para ejecutar el nuevo método. Una parte esencial es una descripción de los procedimientos de control de la calidad que serán usados para monitorear el desempeño en el trabajo de rutina. Una vez que el analista haya sido entrenado y el método incluido en el servicio de rutina, será muy importante monitorear de cerca el desempeño durante el primer mes, identificar las fuentes de los problemas potenciales, mejorar los procedimientos de mantenimiento preventivo, y actualizar a los analistas acerca de cómo manejar mejor la calidad del método.

¿Qué hacer?

Para realizar un estudio de validación de métodos usted necesita entender los experimentos y organizarlos en un plan eficiente. El bosquejo que hemos ofrecido en este capítulo provee un punto de inicio, pero puede y debería ser modificado para sus aplicaciones particulares. Debería definir y escribir su propio plan, ajustarlo con procedimientos que describan la cantidad de muestras y su origen, así como hojas de trabajo para registrar los datos. Es importante observar los datos a medida que son obtenidos para identificar de manera temprana problemas

mientras son realizados los experimentos. Esto generalmente significa tabular y graficar los datos antes del análisis estadístico. ¡Y los problemas de interés son tanto aquellos surgidos de los experimentos como aquellos que están vinculados con el método! Debe sacar conclusiones acerca de la aceptabilidad del desempeño conforme complete cada experimento individual. Por último, la decisión final debería estar basada en el error máximo o total esperado de los experimentos para precisión a largo plazo y para comparación de métodos. Incluso cuando el método sea juzgado como aceptable, el trabajo no está terminado. La implementación requiere del diseño de un procedimiento de Control Estadístico Interno de la Calidad adecuado, de la documentación de los procedimientos estándar de operación, y del entrenamiento eficiente de todos los operadores. La validación de métodos no termina hasta que el método demuestra un desempeño estable bajo condiciones de operación de rutina.

Referencias

1. CLSI EP5-A2. Evaluation of precision performance of quantitative measurement methods. Approved Guideline 2004. Clinical and Laboratory Standards Institute, Wayne PA.
2. CLSI EP9-A2. Method comparison and bias estimation using patient samples. Approved Guideline 2002. Clinical and Laboratory Standards Institute, Wayne PA.
3. CLSI EP15-A2. User verification of performance for precision and trueness. Approved Guideline 2006. Clinical and Laboratory Standards Institute, Wayne PA.

7: ¿Cómo se analizan los datos experimentales?

Antes de explicar los detalles de los experimentos específicos, el Dr. Westgard da una visión general del análisis de datos que es útil y apropiada para los diferentes experimentos. El acercamiento presentado se basa en considerar las “herramientas” para el análisis de datos, más que las estadísticas y las ecuaciones. Estas herramientas están disponibles de manera simple con un formato de calculadores, hojas de cálculo electrónicas y programas de computadora. Se introducen calculadores en línea para proveer herramientas fáciles de usar con este libro de texto.

Objetivos:

- Minimizar su miedo a la estadística.
- Identificar las herramientas y las técnicas necesarias para el análisis de datos.
- Combinar las herramientas con los experimentos y los errores a ser estimados.
- Reconocer la capacidad de las herramientas de cálculo disponibles

Materiales del capítulo:

- VM — **El conjunto de herramientas de análisis de datos**, por James O. Westgard, PhD
- **El conjunto de herramientas de análisis de datos para la validación de métodos**, <http://www.westgard.com/mvtools.html>

Cosas qué hacer:

- Estudiar la lección.
- Practicar usando los calculadores en línea con los datos presentados como ejemplo.
- Revisar las estadísticas presentadas en un informe de validación publicado.

Preguntas de Autoevaluación:

- ¿Qué cálculos realizaría usted con los datos obtenidos a partir del experimento de replicación?
- ¿Qué cálculos realizaría usted con los datos obtenidos a partir del experimento de comparación de métodos?
- ¿Qué gráficos están asociados a los diferentes tipos de análisis?
- Estime la media, desvío estándar y coeficiente de variación a partir del siguiente conjunto de datos:
[203, 202, 204, 201, 197, 200, 198, 196, 206, 198, 196, 192, 205, 195, 207, 198, 201, 195, 202, 195]
- Estime los estadísticos de regresión y el sesgo promedio a partir del siguiente conjunto de datos:

Método de Examen (Y)	Método de Comparación (X)
218	217
161	161
240	244
193	193
290	295
117	118
118	122
203	204
74	74
114	116
245	238
262	260
203	203
218	207
311	304
362	353
332	327
428	423
163	155
268	257

Validación de Métodos: La “Caja de Herramientas” para el Análisis de Datos

James O. Westgard, PhD

Este capítulo en realidad es sobre estadística, pero no puse “estadística” en el título porque mucha gente se desanima tan pronto como ve esa palabra. Otras se ponen incómodas cuando ven las ecuaciones para los cálculos estadísticos. Hasta ahora—tres oraciones en este capítulo—usted podría preguntarse si puede simplemente saltarse el capítulo y evadir el tema. La respuesta es NO. Necesita la estadística para darle sentido a los datos recolectados en los experimentos de validación de métodos.

¡Herramientas, no ecuaciones!

Para reducir los obstáculos mentales que surgen cuando tratamos de entender la estadística, **no hay ninguna ecuación en este capítulo**. En su lugar, vamos a asumir que los cálculos pueden ser realizados fácilmente con la calculadora y la tecnología de informática disponible hoy en día. Su trabajo principal será reconocer *qué* cálculos son útiles para diferentes grupos de datos.

Cuando doy conferencias sobre este tema, comienzo mostrando a la clase un conjunto de herramientas, tales como un martillo, una llave para tuercas, una sierra y un destornillador. Las herramientas de oficina (tales como una engrapadora, tijeras, papel y pluma) también son buenos ejemplos, pero usted está demasiado cómodo con esas herramientas. Quiero que aprenda que puede usar herramientas, incluso si no está muy cómodo con ellas. Así que, consideremos el martillo, la llave para tuercas, la sierra y el destornillador.

- ¿Qué herramienta sería más útil para colgar un cuadro en la pared?
- ¿Qué herramienta usaría para ajustar el armazón de sus lentes de sol?
- ¿Qué herramienta quisiera llevar en Navidad cuando va al bosque para conseguir su árbol?
- ¿Qué herramienta espera tener si su carro tiene una llanta baja?

Usted no tiene que ser un ingeniero, un mecánico o un carpintero para reconocer qué herramienta se ajusta a cada uno de estos trabajos. Cada quien hace uso de estas herramientas para hacer ciertos trabajos básicos. Si bien existen aplicaciones más complicadas que requerirían más habilidad y conocimiento —y

algunas veces herramientas más especializadas— cualquiera es capaz de hacer uso práctico de las herramientas comunes.

¡Podría ser lo mismo con la estadística!

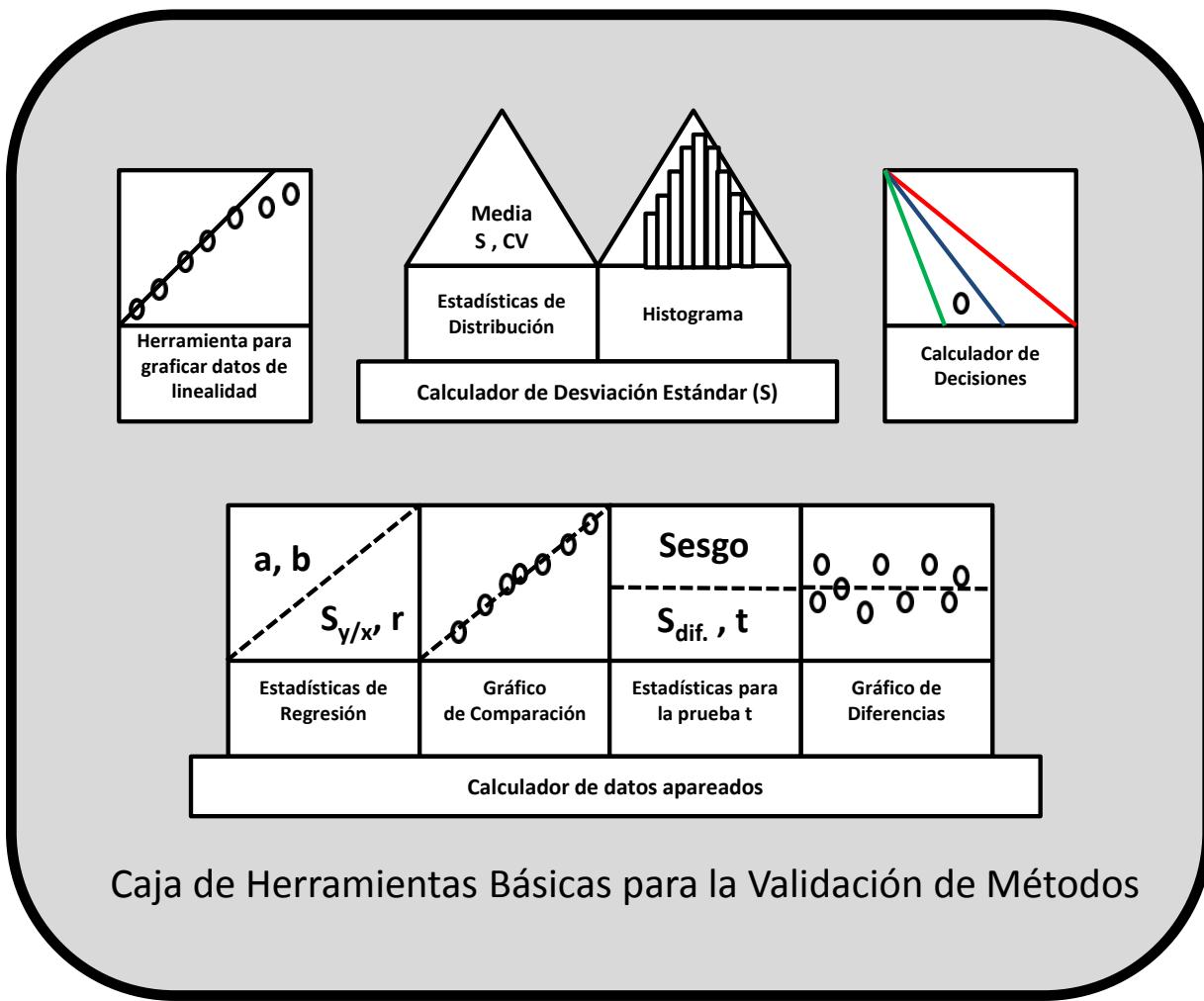
Herramientas recomendadas para el análisis de datos

La estadística sólo es una herramienta para combinar muchos resultados experimentales, es decir datos, y resumir todos estos datos en sólo unos cuantos números. Recuerde que el objetivo de cada experimento es estimar la cantidad de error a partir de los datos recolectados. La clave con la estadística es saber cuál estadístico nos da una información más útil acerca de los errores de interés en los diferentes experimentos.

Antes de intentar estimar estos errores, necesitamos definir el rango analítico útil (o rango reportable) del método para que los experimentos puedan ser planificados apropiadamente y puedan ser recolectados datos válidos. El rango reportable está usualmente definido como el rango en el cual la respuesta analítica del método es lineal con respecto a la concentración del analito que está siendo medida.

Luego comenzamos con el análisis de errores. Primero, queremos saber la imprecisión o el error aleatorio de 20 ó más datos obtenidos a partir de un experimento de replicación. Luego necesitamos estimar el error sistemático a partir de 40 o más datos (para cada uno de los métodos, o sea 40 pares de datos) obtenidos a partir de un experimento de comparación de métodos. Finalmente, necesitamos evaluar desempeño del método en base a los errores que han sido observados. La estadística se usa para hacer estimaciones confiables de los errores a partir de los datos que han sido obtenidos.

Aquí hay una imagen de la caja de herramientas que necesitará para analizar los datos de los experimentos de la validación básica de métodos. La caja de herramientas incluye varios calculadores y herramientas para graficar:



- **Herramienta para graficar datos de linealidad** para mostrar la respuesta del método observado contra las concentraciones relativas o asignadas para una serie de soluciones o especímenes;
- **Calculador de Desvío Estándar (S)** para determinar la estadística de la distribución (media, S, CV) y para mostrar un histograma de la distribución;
- **Calculador de datos apareados** para determinar estadísticas de regresión (pendiente o “a”, intercepción en “y” o “b”, desviación estándar sobre la linea de la regresión lineal o “ $s_{y/x}$ ”, y el coeficiente de correlación, “r”), mostrar los datos en un “gráfico de comparación” (método de examen en “y”, método de comparación en “x”), determinar los estadísticos “t” (sesgo, $S_{dif.}$ valor-t), y mostrar datos en un gráfico de diferencias ($y-x$ contra x);
- **Calculador de Decisiones** para evaluar el desempeño.

Nótese que estas herramientas a menudo incluyen tanto los cálculos como gráficos a partir de los datos. Hay una asociación entre ciertos cálculos y gráficos porque se complementan entre sí al describir y mostrar un conjunto de datos. Por ejemplo, las estadísticas de distribución se utilizan con un histograma para describir y mostrar datos para imprecisión o error aleatorio. Para la inexactitud o el error sistemático, se utilizan estadísticas de regresión con un gráfico de comparación, o estadísticas “t” con un gráfico de diferencias.

Nótese también que existe un orden natural para usar las herramientas, como se sugiere por su ubicación en la caja de herramientas. Aquellas situadas arriba son generalmente empleadas primero, por ejemplo, la herramienta para graficar datos lineales es usada en un primer momento para establecer o verificar el rango reportable del método, después la próxima herramienta a utilizar es el Calculador de Desvío Estándar para estimar la imprecisión o el error aleatorio, cuya aceptabilidad puede ser evaluada utilizando la calculadora de decisiones. Después de estos pasos, la calculadora de datos apareados será utilizada para estimar la inexactitud del método y usar de nuevo el Calculador de Decisiones para evaluar el desempeño general del método.

Dónde conseguir las herramientas

Estas herramientas de cálculo podrían ser obtenidas empleando calculadoras de mano (por ejemplo, Texas Instruments), de hojas de cálculo (por ejemplo, Excel), software de estadística de uso general (Minitab, SAS, SPSS), software especializado de validación de métodos hecho para aplicaciones del laboratorio, y también de calculadores en línea en la página web de Westgard. Muchas de estas fuentes también van a ofrecer gráficos apropiados, o puede usted construirlos manualmente usando papel para graficar. La Carta de Decisión de Método puede ser construida fácilmente a mano con papel para graficar.

El capítulo 8 provee más información acerca de los cálculos. Los capítulos posteriores nos ofrecen una discusión más detallada acerca del uso e interpretación de las estadísticas con experimentos individuales. Por ahora, vamos a enfocarnos en un contexto global —qué herramientas son apropiadas para los diferentes experimentos de validación de métodos.

Cuándo usar cada herramienta

Dado un conjunto de datos experimentales, necesita reconocer qué herramienta es correcta cada trabajo específico. Aquí hay unas guías generales:

- **El error aleatorio (EA)** casi siempre se estima calculando la desviación estándar (S). El experimento por sí mismo determina cuáles son los factores

que contribuyen al estimado, por ejemplo, el experimento de replicación limita los errores aleatorios sólo al método que se está evaluando, mientras que el experimento de comparación de métodos puede proveer un estimado de los errores aleatorios entre los métodos, lo que depende de la variación observada para el método en evaluación (examen), y para el método de comparación.

- **El error sistemático (ES)** está relacionado en cierta manera con el cálculo de una media o promedio. Ésta podría ser la diferencia promedio entre muestras apareadas en un estudio de comparación de métodos, o la diferencia entre las medias obtenidas a partir de cada método, o una representación de la relación promedio como se da por la línea de mejor ajuste a través de datos de comparación de métodos.
- Recuerde que una decisión sobre la **aceptabilidad** del desempeño de un método implica evaluar si los errores observados afectarán la utilidad clínica del examen. La estadística provee el mejor estimado del tamaño de los errores [1]. Usted debe evaluar si esos errores afectarán la utilidad clínica del examen [2]. Puede hacer esto definiendo un requisito de la calidad en la forma de un error total permitido, TEa, tal como se define en los criterios de los esquemas de evaluación de la competencia de CLIA para desempeño aceptable. Una herramienta gráfica simple llamada la Carta de Decisión de Método puede ser usada para ayudarle a evaluar el desempeño del método.

Herramientas de ejemplo para uso educacional

Las calculadoras en línea para uso educacional están disponibles en la página web de Westgard <http://www.westgard.com/mvtools.html>. Estas herramientas web deberían ser útiles para trabajar con los conjuntos de datos empleados para los ejemplos y para el conjunto de datos presentados como problemas para resolver. Sin embargo, no están diseñados para responder a todas sus necesidades de análisis de datos para los estudios de validación de métodos. También se recomienda que adquiera sus propias herramientas de cálculo, ya sea un programa de estadística general, un programa de validación de métodos especializado, o una hoja de cálculo electrónica.

La herramienta para graficar datos de linealidad es usada con los datos obtenidos en el experimento de linealidad, donde el propósito es evaluar el rango analítico, a lo largo del cual serían informados los resultados de pacientes. La respuesta del método es graficada en el eje “y” contra la concentración relativa o valores asignados de las muestras o especímenes en el eje “x”. El “rango reportable” es estimado generalmente como el rango lineal de trabajo del método analítico.

El calculador de desvío estándar se utiliza para los datos obtenidos en el experimento de replicación, donde el objetivo es estimar el error aleatorio o la imprecisión del método en base a mediciones repetidas del mismo material. La estadística que debería ser calculada es la media, la S y el CV. También asegúrese de registrar el número de mediciones usadas en los cálculos.

- La media, o el promedio del grupo de resultados, describe la locación central de las mediciones.
- La desviación estándar “S” describe la distribución de resultados esperada, esto es, se espera que el 66% de los resultados estén dentro de la media $\pm 1 S$, el 95% dentro de la media $\pm 2 S$, y el 99.7% dentro de la media $\pm 3 S$.
- El CV, o coeficiente de variación, es igual a la desviación estándar “S” dividida por la media, multiplicada por 100 para expresarse en porcentaje.
- El histograma muestra la distribución de resultados. Idealmente, la distribución debería ser Gaussiana, o “normal”.

El calculador de datos apareados podría ser usado con los pares de resultados de cada espécimen analizado por el método en evaluación (examen) y el método de comparación en el experimento de comparación de métodos. Esta es la parte más complicada del análisis estadístico y requiere el mayor cuidado y atención. La estadística de regresión lineal podría ser usada con un gráfico de comparación, usar la estadística de la prueba “t” junto con un gráfico de diferencias.

Las estadísticas de regresión que deberían ser calculadas son la pendiente (b) y la intercepción en “y” de la línea (a), la desviación estándar de los puntos alrededor de esa línea (s_{yx}), y el coeficiente de correlación (r el coeficiente de correlación del momento del producto Pearson). También podría ver la pendiente designada como “m”, la intercepción en “y” como “b” y la desviación estándar como $s_{residuos}$, respectivamente. Se incluye el coeficiente de correlación para ayudarle a decidir si las estadísticas de regresión lineal o las estadísticas “t” nos van a dar los estimados más confiables del error sistemático.

- La pendiente describe el ángulo de la línea que provee el mejor ajuste entre los resultados generados por el método en evaluación y los obtenidos por el método de comparación. Una pendiente perfecta sería 1.00. Las desviaciones de 1.00 son una indicación del error sistemático proporcional [1].
- La intercepción en “y” describe dónde se intersecta la línea de mejor ajuste con el eje “y”. Idealmente, la intercepción en “y” debería ser 0.0. Las desviaciones de 0.0 son indicadores de un error sistemático constante.

- El término s_{xy} describe la dispersión de los datos alrededor de la línea de mejor ajuste. Brinda una estimación del error aleatorio entre los métodos que incluye tanto la imprecisión del método de examen y como la del método de método de comparación, así como posibles efectos matriz que varían de un espécimen a otro. Nunca será cero porque tanto el método de examen como el método de comparación tienen alguna imprecisión.
- El coeficiente de correlación el grado de acuerdo entre los resultados de los dos métodos (el método en evaluación y el de comparación). Un r de +1.00 indica una correlación perfecta, esto es, todos los puntos caen perfectamente en una línea que muestra los valores del método de examen contra los valores del método de comparación. Los valores menores a 1.00 indican que hay dispersión en los datos alrededor de la línea de mejor ajuste. Entre más bajo el valor de r , más dispersos los datos. El uso principal de r es ayudarle a evaluar la fiabilidad de los cálculos de regresión lineal — **el r nunca debe ser usado como un indicador de la aceptabilidad del método**. Cuando r es 0.99 o mayor, los cálculos de regresión lineal proveerán estimaciones confiables de los errores. Cuando r es menor a 0.975, es mejor usar cálculos de datos apareados (prueba “t”) o técnicas de regresión alternativas (y más complicadas) tal como la regresión de Deming o de Passing-Bablock [4,5].
- Se debería utilizar un gráfico de comparación para mostrar los datos del experimento de comparación de métodos (graficando los valores del método de comparación en el eje “x” y los valores del método de examen en el eje “y”). Este gráfico es usado entonces para inspeccionar visualmente los datos, para identificar posibles valores atípicos y evaluar el rango lineal de acuerdo [1].

Las estadísticas de la prueba “t” de interés son el sesgo, el desvío estándar de las diferencias (S_{dif}), y finalmente, lo que se conoce como valor “t” que también requiere conocimiento del número de mediciones de muestras apareadas. De nuevo, asegúrese de llevar registro del número de mediciones, el cual para el experimento de comparación de métodos es el número de especímenes de pacientes comparados.

- El sesgo es la diferencia entre los promedios de los dos métodos, que también es la diferencia promedio de todos los especímenes analizados por los dos métodos. Nos da un estimado del error sistemático o de la diferencia promedio que se espera entre los dos métodos — entre más pequeño el sesgo, más pequeño el error sistemático, mejor el grado de acuerdo.

- El desvío estándar de las diferencias (S_{dif}) provee un estimado del error aleatorio entre los métodos. Nunca será cero porque tanto el método de examen como el método de comparación tienen algo de imprecisión.
- El valor “t” por sí mismo es un indicador de si se han recolectado suficientes mediciones de muestras apareadas para saber si el sesgo observado es real, o estadísticamente significativo. Por regla general, en un experimento de comparación de métodos donde se ha comparado un mínimo de 40 especímenes de pacientes (procesados en ambos métodos), si es mayor que 2.0, implica que los datos son suficientes para concluir que existe un sesgo. Es importante recordar que para evaluar la aceptabilidad del método debemos considerar el tamaño del sesgo y no del valor “t”.
- Un gráfico de diferencias debería ser usado para mostrar las diferencias entre los resultados apareados, graficando la diferencia entre los resultados del método de examen menos los valores obtenidos por el método de comparación en el eje “y” para cada espécimen contra los resultados obtenidos a partir del método de comparación en el eje “x”. Los gráficos de diferencias son populares hoy en día por su simplicidad, sin embargo, su uso e interpretación no son tan simples cuando se quiere tomar una decisión cuantitativa objetiva acerca del desempeño.

El calculador de decisiones es utilizado para mostrar las estimaciones de los errores aleatorios y sistemáticos y evaluar el desempeño del método. Por consiguiente, esta gráfica depende de las estimaciones de los errores que son obtenidas a partir de otros cálculos estadísticos. Resumiendo, la gráfica es trazada en base al requisito de la calidad definido para el método y muestra la inexactitud permitida en el eje “y” contra la imprecisión permitida en el eje “x”. La imprecisión y la inexactitud observadas del método son luego graficadas para mostrar el “punto operativo” del método (la coordenada “y” es la estimación de la inexactitud o el ES, la coordenada “x” es la estimación de la imprecisión o el EA). La posición de este punto operativo es interpretada en relación a las líneas que definen áreas de desempeño “inaceptable”, “pobre”, “marginal”, “bueno”, “excelente” y “clase mundial” (Seis Sigma). Vea los archivos en PDF en la página web de Westgard para más detalles. Discutiremos el Cuadro de Decisión del Método en detalle en el capítulo 16.

Referencias

1. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method comparison studies. Clin Chem 1973;19:49-57. See below for PDF files.
2. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974;20:825-33.
3. Westgard JO. A method evaluation decision chart (MEDx Chart) for judging method performance. Clin Lab Science. 1995;8:277-83.
4. Stockl D, Dewitte K, Thienpont M. Validity of linear regression in method comparison studies: limited by the statistical model or the quality of the analytical data? Clin Chem 1998;44:2340-6.
5. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. Clin Chem 1979;25:432-8.
6. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;307-10.
7. Hyltoft-Petersen P, Stockl D, Blaabjerg O, Pedersen B, Birkemose E, Thienpont L, Flensted Lassen J, Kjeldsen J. Graphical interpretation of analytical data from comparison of a field method with a reference method by use of difference plots [opinion]. Clin Chem 1997;43:2039-46.

Referencias en línea

The Method Validation data analysis tool kit
<http://www.westgard.com/mvtools.html>

Use and interpretation of common statistical tests in method comparison studies. <http://www.westgard.com/method1.htm>



¡Otro problema de perspectiva!

CHINA, by Philip Galle, Antwerp, 1595.

China aparece diferente en este mapa porque el Norte está a la derecha, no arriba. No obstante la Gran Muralla se muestra con claridad. Su perspectiva sobre las estadísticas es crítica para su uso correcto y aplicación. Si usted puede recordar que las estadísticas son herramientas para estimar el tamaño de los errores, entonces podrá tener bien clara la información que es importante.

8: ¿Cómo se calculan estas estadísticas?

Este capítulo brinda las ecuaciones que fueron evitadas tan cuidadosamente en el capítulo anterior de herramientas para el análisis de datos. Es muy importante que usted entienda las pruebas de significancia estadística y sus limitaciones al momento de evaluar la aceptabilidad de un método y la verificar se obtenga un desempeño semejante al declarado por el fabricante en sus especificaciones de desempeño. También es importante comprender el uso y aplicación de las estadísticas que puede encontrar en los programas de computación específicos disponibles. Se necesita acceder a los cálculos correctos de aquellos programas y algunas veces es necesario mirar las ecuaciones para confirmar que está obteniendo los cálculos correctos.

Objetivos:

- Aprender los términos estadísticos que se encuentran comúnmente en los informes de validación de métodos y las especificaciones de desempeño del método declaradas por los fabricantes.
- Revisar las ecuaciones utilizadas para calcular esas estadísticas.
- Reconocer aspectos críticos del cálculo y la interpretación de los estadísticos de significancia “t” y “F” y su capacidad para verificar las especificaciones de desempeño del método declaradas por los fabricantes.
- Identificar las estadísticas incluidas comúnmente en los análisis de regresión, incluyendo la pendiente, la intersección-Y, el error estándar de la línea de regresión, la desviación estándar de la pendiente, la desviación estándar de la intersección, y el coeficiente de correlación.

Materiales del capítulo:

- VM — Los Cálculos Estadísticos por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar los materiales.
- Identificar un programa de cómputo disponible para su trabajo de rutina.
- Encontrar todas las estadísticas en su programa de cómputo.

Preguntas de Autoevaluación:

- En un estudio de validación de métodos de colesterol, el estadístico-t obtenido a partir de la comparación de métodos arroja los siguientes datos:
 $Sesgo = 1.74 \text{ mg/dL}$, $S_{dif} = 5.90 \text{ mg/dL}$, $N = 81$, y $t = 2.65$. ¿Cuál es el valor crítico de t para $p = 0.05$? ¿Es el sesgo observado estadísticamente significativo? ¿Es el sesgo observado clínicamente significativo?
- En un estudio de validación de métodos, el método en evaluación tiene una desviación estándar de 4.0 mg/dL y el método de comparación tiene una desviación estándar de 5.0 mg/dL, ambos valores fueron estimados a partir de 21 mediciones efectuadas sobre el mismo material de control cuya concentración aproximada es de 200 mg/dL. ¿Cuál es el valor calculado para el estadístico “F”? ¿Es la diferencia de desempeño a nivel de precisión estadísticamente significativa? ¿Es la diferencia de desempeño a nivel de precisión clínicamente significativa?
- Para verificar las especificaciones de precisión declaradas por el fabricante se calculó el estadístico-F para comparar la desviación estándar de 5.0 mg/dL obtenida en un experimento de replicación para el método evaluado a partir de un $N= 31$ con la desviación estándar declarada por el fabricante de 4.0 mg/dL obtenida también de un experimento de replicación con $N= 31$. ¿Ha verificado usted las especificaciones del fabricante para precisión?
- Para verificar las especificaciones de precisión declaradas por el fabricante se calculó el estadístico-F para comparar la desviación estándar de 5.0 mg/dL obtenida en un experimento de replicación para el método evaluado a partir de un $N= 11$ con la desviación estándar declarada por el fabricante de 3.0 mg/dL obtenida también de un experimento de replicación con $N= 11$. ¿Ha verificado usted las especificaciones del fabricante para precisión?
- Para verificar las especificaciones del fabricante para exactitud se trabajó con el estadístico-t, se obtuvo un sesgo de 1.5 mg/dL y un valor de “t” de 2.5 para un estudio donde se procesaron 40 muestras de pacientes por el método a evaluar y el mismo método de comparación empleado por el fabricante al momento de establecer sus especificaciones. El fabricante declaró que no existía sesgo entre ambos métodos. ¿Ha verificado usted las especificaciones declaradas por el fabricante para exactitud?

Validación de Métodos: Los Cálculos Estadísticos

James O. Westgard, PhD

La estadística provee una herramienta para lograr que un conjunto de resultados de experimentos individuales puedan ser combinados y resumidos con unos pocos números. Esto debería ser de gran ayuda para clarificar el significado de los resultados experimentales. Sin embargo, la estadística a menudo tiene el efecto contrario, causando más bien confusión que clarificación. Esto ocurre usualmente cuando no se entiende bien el propósito del análisis estadístico. En la validación de métodos, el objetivo debería ser estimar los errores analíticos a partir de los resultados de los experimentos de validación. La estadística debería proporcionar información sobre el tipo y magnitud de los errores analíticos: de otro modo tendrían poco valor para interpretar los resultados de los estudios de validación y para evaluar la aceptabilidad del desempeño del método.

Una estadística es simplemente un número. Es un número que describe una serie de otros números a partir de los cuales ha sido calculada. La media (\bar{X}) de un conjunto de números se calcula en los experimentos de replicación, recuperación e interferencia. La desviación estándar (S) se calcula en el experimento de replicación. Estas estadísticas se calculan a partir de un número limitado de observaciones, las cuales son un subconjunto o *muestra estadística* de la población de interés. Proveen estimados que son usados para inferir algo acerca de la *población* como un todo; aunque estos estimados no son necesariamente idénticos a los números o *parámetros* que describen a la población. Por ejemplo, \bar{X} y S son estadísticas utilizadas para inferir los valores de parámetros con ser, la media verdadera, y σ , la verdadera desviación estándar. Sin embargo, \bar{X} y S podrían no tener los mismo valores que μ y σ , especialmente cuando el número de mediciones es pequeño. Se pueden hacer inferencias y conclusiones de fiar sólo cuando el trabajo experimental es el adecuado y los análisis estadísticos y la interpretación se hace correctamente.

El uso correcto y la interpretación de la estadística es esencial si los experimentos de validación se realizan para tomar decisiones objetivas sobre el desempeño del método. La estadística utilizada comúnmente en la validación de métodos incluye la media, la desviación estándar, el coeficiente de variación, el error estándar de la media, el estadístico “t” obtenido a partir de datos apareados, el estadístico “F”, el análisis de regresión lineal y el coeficiente de correlación.

Media, Desviación Estándar, Coeficiente de Variación

La media (\bar{X}) es una medida de tendencia central o localización de un conjunto de valores. La *desviación estándar* (S) es una medida de *dispersión* o propagación de un conjunto de mediciones alrededor de la media. El *coeficiente de variación* (CV) expresa la desviación estándar como un porcentaje de la media. Las ecuaciones son las siguientes:

$$\bar{x} = \frac{\sum x_i}{N}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}}$$

$$CV = \frac{s (100\%)}{\bar{x}}$$

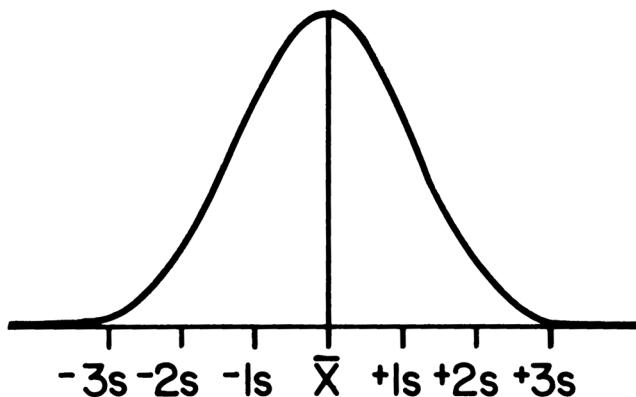
donde x_i es una observación individual, \bar{X} es la media de la serie de observaciones, y N es el número total de observaciones en la serie.

El denominador de la segunda ecuación utiliza $N - 1$ porque esta es una ecuación para la estadística S, más que una ecuación para el parámetro σ , en cuyo caso se utilizaría N. La estadística S da el mejor estimado del parámetro σ cuando se usa $N - 1$ y subestimaría σ si se utilizara N. Por supuesto, cuando N es grande, hace poca diferencia si se usa N o $N - 1$.

$N - 1$ también representa el número de *grados de libertad* (GL). Este es un concepto difícil. Los grados de libertad son definidos como “el número de comparaciones independientes que pueden ser realizadas entre N observaciones”. Esto también podría ser considerado como el número de observaciones o mediciones en una serie menos el número de restricciones en la serie. Por ejemplo, si hubiera tres mediciones, sólo serían necesarias tres ecuaciones simultáneas para identificar los tres valores. Si se calculará la media, entonces sólo habría $N - 1$ ecuaciones independientes sobrantes, o $N - 1$ comparaciones independientes, o $N - 1$ grados de libertad para una desviación estándar. Esto se vuelve más obvio si $N=1$, en cuyo caso una única observación podría ser tomada como un estimado de la media, pero no sería posible estimar una desviación estándar. Sólo hay una pieza de información y no puede ser usada para estimar otras dos cantidades.

Si se utiliza la desviación estándar para describir un conjunto de datos, se asume que la *distribución* de los datos se ajusta a una *curva Gaussiana* o a una *curva normal*. Esta es una distribución simétrica con forma de campana que puede ser descrita matemáticamente por una ecuación específica en la cual la media y la desviación estándar son las variables. La estadística que asume

una distribución Gaussiana de una variable se llama *estadística paramétrica*. Aquella que no hace ninguna asunción acerca de la naturaleza de la distribución de una variable se llama *estadística no paramétrica*. Toda la estadística que se discute aquí es estadística paramétrica.



Esta figura muestra una curva Gaussiana. En una curva Gaussiana, 68.2 por ciento de las observaciones estarán dentro de $\pm 1.0s$ de la media, 95.5 por ciento dentro de $\pm 2.0s$ de la media, y 99.7 por ciento dentro de $\pm 3.0s$ de la media. Es común hablar sobre un intervalo de confianza o rango de confianza de un 95 por ciento y estimar esto por la media $\pm 2.0s$, aunque el múltiplo exacto debería ser 1.96 para un intervalo de 95.0 por ciento.

Error Estándar de la Media

La dispersión de los valores de la media está dada por una estadística llamada el *error estándar de la media* (S_x), que se calcula a partir de la siguiente ecuación:

$$S_x = \frac{S}{\sqrt{N}}$$

donde “S” es la desviación estándar que representa la dispersión de mediciones individuales como un estimado de mediciones replicadas y N es el número de mediciones incluidas en la media.

En los experimentos de validación destinados a calcular los errores sistemáticos, tales como los experimentos de recuperación e interferencia, a menudo es útil promediar los resultados de varias mediciones o muestras. Esto reduce los efectos del error aleatorio del método y permite una evaluación de los errores sistemáticos pequeños en un método que tiene un error aleatorio relativamente grande. La media verdadera de las mediciones estará contenida por un inter-

valo de confianza alrededor de la media experimental, como se describe en la ecuación siguiente:

$$\mu = \bar{x} \pm t \frac{s}{\sqrt{N}}$$

El ancho del intervalo se calcula del error estándar de la media “ S_x ” ($s/(N)^{1/2}$) y un múltiplo (“ t ”). Éste múltiplo debe ser apropiado para los grados de libertad, que son $N - 1$, y el nivel del intervalo de confianza deseado, usualmente 90%, 95% o 99%. El valor numérico del múltiplo t se obtiene de una tabla estadística, que usualmente enumera t como una función de los grados de libertad y una *probabilidad* (p). Los valores t para $p=0.10$ se utilizan para calcular un intervalo de confianza de 90%, aquellos para $p=0.05$ son usados para un intervalo de confianza de 95%, y aquellos para $p=0.01$ son usados para un intervalo de confianza de 99%.

Se muestra en la página 95 una tabla ejemplo de valores “ t ”. Observe que los valores “ t ” para $p=0.05$ se acercan a 1.96 conforme el grado de libertad se acerca a infinito. En general, un múltiplo de 2 puede ser utilizado para calcular un intervalo de confianza del 95% cuando $GL=20$ ó más, y el intervalo será correcto dentro de un cuatro por ciento. Ésta es la base del uso general de $+ 2s$ como un estimado del 95% del intervalo de confianza. Note que, a medida que GL se hace más pequeño, especialmente menor de diez, “ t ” se hace mucho mayor. He aquí por qué es importante obtener el valor “ t ” correcto para experimentos tales como los de recuperación e interferencia, donde GL o $N-1$ tienden a ser pequeños.

Aquí hay un ejemplo del uso de la ecuación para calcular un estimado del intervalo de confianza del error sistemático. Use los siguientes datos de un experimento de recuperación: 94 por ciento, 96 por ciento, 95 por ciento, 96 porcentaje, 99 por ciento, 97 por ciento, 100 por ciento, 96 por ciento y 98 por ciento. De estos nueve valores, la media de recuperación (X) es 96.8 por ciento y la desviación estándar (S) expresada en porcentaje (CV) es 1.9 por ciento. Por lo tanto la verdadera media de recuperación (μ) estimada con 95% de confianza es

$$\mu = 96.8 \pm 2.31 \frac{1.9}{\sqrt{9}}$$

donde 2.31 es el valor “ t ” para $GL=8$ ($N - 1$) y $p=0.05$, como se obtiene de la tabla en la página 95. El intervalo que va desde 95.3% hasta 98.3% tiene un 95% de probabilidad de contener la verdadera media de recuperación. Sólo hay

un 5% de probabilidad de que la verdadera media de recuperación esté fuera de este intervalo.

Se pueden realizar cálculos similares para el experimento de interferencias. Por ejemplo, los siguientes pares de observaciones (muestra blanco, muestra para examen) muestran el efecto de añadir una interferencia sospechada: 20, 23; 14, 20; 15, 20; 12, 15; 32, 35; 22, 26; 40, 42; 18, 23; y 16, 22. Las interferencias observadas de las muestras individuales son 3, 6, 5, 3, 3, 4, 2, 5 y 6. La media observada de la interferencia (\bar{X}) es 4.1 y la desviación estándar (S) es 1.5. La verdadera media de la interferencia (μ) es estimada con 95% de confianza como:

$$\mu = 4.1 \pm 2.31 \frac{1.5}{\sqrt{9}}$$

donde 2.31 es nuevamente el valor de t para GL=8 y p=0.05. El intervalo que va desde 2.9 hasta 5.3 tiene un 95% de probabilidad de contener la verdadera media de la interferencia. Sólo existe un 5% de probabilidad de que la verdadera media de la interferencia esté fuera de este intervalo.

La utilidad de establecer un intervalo de confianza como estimado del error sistemático debería volverse visible al comparar los resultados de estos dos experimentos de recuperación:

- (A) (\bar{X}) es 96.8%, el intervalo de confianza del 95% es de 95.3% a 98.3%;
- (B) (\bar{X}) es 96.8%, el intervalo de confianza del 95% es de 85.3% a 108.3%.

En el experimento A, los datos son suficientes para concluir que existe un error sistemático proporcional debido a que el intervalo de confianza no se superpone al valor ideal de 100% de recuperación. En el experimento B, los datos no son suficientes para apoyar una conclusión de la existencia de un error proporcional debido a que el intervalo de confianza se superpone al valor ideal de 100%.

Las pruebas de significación y los intervalos de confianza

Los exámenes estadísticos tales como el estadístico “t” y el estadístico “F” son usados a menudo para determinar si existe diferencia entre dos cantidades que son estimados de los parámetros de desempeño. Estos exámenes son llamados *pruebas de significación* y miden si los datos experimentales son adecuados para apoyar una conclusión sobre la presencia de una diferencia real. La hipótesis

examinada se llama *hipótesis nula* que dice que no hay diferencia entre las dos cantidades. Cuando el estadístico (t o F) es grande, se desaprueba la hipótesis nula. La conclusión es que la diferencia es *estadísticamente significativa*. En términos prácticos, esto significa que se ha observado una diferencia real. Cuando el estadístico es pequeño, la conclusión es que la hipótesis nula se mantiene y *no hay diferencia estadísticamente significativa* entre las dos cantidades. No se ha observado una diferencia real.

Los pruebas de significación proveen información similar a la disponible a partir de los intervalos de confianza, y las similitudes son útiles para entenderlas. Cuando el intervalo de confianza del estimado de una cantidad no se superpone a un valor de interés especificado, dígase cero, esto significa que el estimado no incluye ese valor especificado y por tanto es diferente de él. Esto es análogo a una prueba de significación, que da un resultado estadístico grande, indicando así una diferencia estadísticamente significativa. Cuando el intervalo de confianza se superpone al valor de interés especificado, la situación es similar a la observación obtenida a partir de la prueba de significación que indica que no hay una diferencia estadísticamente significativa.

Estadístico t para datos apareados

Un estadístico “t” puede ser usado para medir dos medias y determinar si existe diferencia entre ellas. Existen formas del estadístico “t” tanto para datos apareados como no apareados. Esto se refiere a si las dos medias comparadas vienen de las mismas muestras estadísticas o de diferentes muestras estadísticas. Por ejemplo, el estadístico “t” para datos apareados es utilizado cuando hay pares de mediciones de un conjunto de muestras tales como en el experimento de comparación de métodos en el que cada muestra es analizada por el método analítico evaluado y el de comparación. La forma no apareada se utiliza cuando se examina la diferencia entre medias de dos conjuntos separados de muestras, tales como la media de los valores de referencia para mujeres contra la media para hombres. Se describen aquí sólo los cálculos de la forma apareada del estadístico “t”.

Se calculan tres estadísticas. El sesgo es la diferencia entre dos medias:

$$\text{Sesgo} = \bar{Y} - \bar{X}$$

donde la \bar{Y} y la \bar{X} son valores promedio. La *desviación estándar de las diferencias* (S_{dif}) está dada por:

$$S_{\text{dif}} = \sqrt{\frac{\sum[(y_i - x_i) - \text{bias}]^2}{N - 1}}$$

Finalmente, el valor “t” es calculado a partir del sesgo, la S_{dif} y N, utilizando la ecuación:

$$t = \frac{\text{Sesgo}}{S_{\text{dif}} / \sqrt{N}}$$

Esta ecuación muestra la naturaleza del valor “t”. Es una proporción de dos términos, uno que representa una diferencia o error sistemático (Sesgo) y otro que representa un error aleatorio ($S_{\text{dif}} / (\sqrt{N})^{1/2}$); en este caso tiene la forma de un error estándar de la media debido a que se están midiendo los valores de la media. El valor de “t” expresa la magnitud del error sistemático en múltiplos del error aleatorio. Por ejemplo, un valor “t” de seis indicaría que el término de error sistemático es seis veces mayor que el término de error aleatorio. Esta cantidad de error sistemático es mucho mayor que la cantidad que podría ser observable sólo con la incertidumbre de los datos experimentales. No se esperan proporciones mayores a dos o tres. Si el valor de la proporción es muy grande, es probable que los datos experimentales demuestren una diferencia real entre los valores promedio, o la presencia de un error sistemático.

GL	Intervalos o pruebas de dos lados		
	p = 0.10	p = 0.05	p = 0.01
1	6.31	12.71	63.66
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.90	2.36	3.50
8	1.86	2.3	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17
12	1.78	2.18	3.06
14	1.76	2.14	2.98
16	1.75	2.12	2.92
18	1.73	2.10	2.88
20	1.72	2.09	2.84
30	1.70	2.04	2.75
40	1.68	2.02	2.70
∞	1.64	1.96	2.58

Tabla “t”. Valores críticos de “t” para probabilidades (p) y grados de libertad (GL) seleccionados.

Una interpretación más exacta del estadístico “t” se realiza comparando el valor “t” observado de la primera ecuación con un valor crítico de “t” obtenido de una tabla estadística. La hipótesis examinada es que no hay diferencia entre los dos valores de la media (hipótesis nula).

Si el valor “t” observado es mayor que el valor crítico de “t” encontrado en la tabla, se rechaza la hipótesis nula, demostrando que hay una diferencia entre los dos valores de la media, o que se ha observado un error sistemático. Usualmente, se declara que la diferencia o el error sistemático es estadísticamente significativo. Esto significa que la diferencia observada es mayor que la esperada debido a la incertidumbre de los datos experimentales o al error aleatorio en las mediciones. La conclusión es que los datos experimentales son suficientes para apoyar una afirmación de que existe error sistemático.

Si el valor “t” observado es menor que el valor crítico de “t” de la tabla, no se rechaza la hipótesis nula. No existe diferencia entre los dos valores de las medias. La diferencia observada o error sistemático no es estadísticamente significativo. Esto quiere decir que los datos experimentales no apoyan una conclusión de que existe error sistemático.

El valor crítico de “t” es tabulado usualmente como una función de probabilidad (p) y de grados de libertad (GL). Es común seleccionar una probabilidad de 0.05 o 0.01 para interpretación de exámenes. Para una interpretación con $p=0.05$, sólo habría una chance del cinco por ciento o probabilidad de 0.05 de que el valor de “t” observado exceda el valor crítico de “t” debido a errores aleatorios en las mediciones. Esto significa que hay un 95% de confianza de que el error sistemático sea real. Para una $p=0.01$, habría sólo un chance del 1% de que la observación responda a errores aleatorios, por lo que hay 99% de confianza de que el error sistemático es real. Los grados de libertad en el estadístico “t” son iguales a $N - 1$. Para un experimento de comparación de métodos con 41 muestras de pacientes, $GL=40$. El valor crítico de “t” para $p=0.05$ y $GL=40$ sería 2.02, como se muestra en la tabla. Para valores observados de “t” mayores a este (2.02), los datos experimentales demuestran que existe un error sistemático. Para valores observados de “t” menores a este (2.02), los datos experimentales no demuestran la presencia de un error sistemático.

Note cuidadosamente que la interpretación no dice *nada* acerca de la aceptabilidad del desempeño del método, solo dice si hay error sistemático presente. La aceptabilidad dependerá del tamaño del error sistemático, no del valor de “t”. Hay situaciones en las que el valor de “t” es grande debido a que el error aleatorio en las mediciones es pequeño o debido a que el número de mediciones es grande. Esto se ve más fácilmente cuando se reordena nuestra ecuación previa:

$$t = \text{Sesgo} \frac{\sqrt{N}}{S_{\text{dif}}}$$

Si el sesgo es constante (y no es cero), “t” incrementará conforme incremente N o disminuya S_{dif} . Por otro lado, cuando S_{dif} sea muy grande, o cuando N sea muy pequeña, “t” tenderá a ser pequeño. Esto causa dificultades al interpretar el valor de “t”, una dificultad que es análoga a la interpretación del valor del pH sanguíneo. El pH sanguíneo es una función de la proporción entre el bicarbonato y el ácido carbónico (o pCO_2). La información acerca de estos componentes individuales es necesaria para una correcta interpretación del valor de pH. De manera similar, la información acerca del sesgo, la S_{dif} y N debe ser considerada para una interpretación correcta del examen.

La aceptabilidad del desempeño se debería juzgar con base en los estimados del error — sesgo y S_{dif} . Los valores ideales para estas estadísticas serían cero. El valor “t” podría abarcar un amplio rango de valores conforme el sesgo y la S_{dif} se acercan a sus valores ideales. Debido a esto, el valor “t” no provee un criterio fiable para juzgar la aceptabilidad, aunque algunas veces se usa de esta manera. Es comúnmente erróneo interpretar un valor “t” pequeño como un indicador de desempeño aceptable, o un valor de “t” grande como un indicador de un desempeño inaceptable. **La única conclusión que debería obtenerse del valor “t” es si se ha demostrado o no que existe error sistemático.**

Estadístico “F”

En estudios de validación de métodos, el *estadístico “F”* se utiliza en ocasiones para comparar la varianza del método analítico evaluado con la varianza del método de comparación. La *varianza* es simplemente el cuadrado de la desviación estándar. Mientras que el estadístico t dice si la diferencia entre dos valores promedio es estadísticamente significativa, el **estadístico “F” dice si la diferencia entre las varianzas es estadísticamente significativa**. Resumiendo, el estadístico “t” se utiliza para error sistemático o inexactitud, y el estadístico “F” se utiliza para el error aleatorio o imprecisión.

Para obtener el estadístico “F”, se elevan al cuadrado las desviaciones estándar del método analítico a evaluar y la del método de comparación, y la varianza más grande es dividida por la varianza más pequeña, como se muestra abajo:

$$F = \frac{(s_1)^2}{(s_2)^2}$$

donde S_1 es la S más grande (o el método menos preciso) y S_2 es la S más pequeña (o el método más preciso). El estadístico “F” se interpreta comparando el valor calculado de “F” con un valor crítico de “F”, que se obtiene de una tabla estadística. La hipótesis nula examinada es que no hay diferencia entre las varianzas de los dos métodos. La hipótesis nula se rechaza cuando el valor observado de “F” es mayor que el valor crítico de “F”, y en este punto, se dice que la diferencia entre las varianzas o los errores aleatorios es estadísticamente significativa.

El significado práctico es simplemente que los datos son suficientes para mostrar que el método en el numerador de la ecuación del estadístico “F” tiene un error aleatorio mayor que el método en el denominador. Cuando el valor observado de “F” es menor que el valor crítico de “F”, la hipótesis nula no puede ser rechazada. No hay diferencia entre las varianzas o los errores aleatorios de los dos métodos.

Por ejemplo, dado un método analítico a evaluar con $S=5.0$ ($N=21$) y un método comparativo con $S=4.0$ ($N=31$), el valor calculado de “F” es 1.56 ($F=25/16$). La tabla abajo (*tabla F*, siguiente página) da una lista abreviada de valores críticos de “F” para $p=0.05$. Los encabezados de columna dan los grados de libertad en el numerador y los encabezados de fila dan los grados de libertad en el denominador. En ambos casos, GL es igual a $N-1$, donde N es el número de mediciones en los respectivos experimentos de replicación. Para este ejemplo, el valor crítico de “F” es 1.93. El valor observado es menor que el valor crítico, lo que significa que no hay diferencia entre los errores aleatorios de los dos métodos (con $p=0.05$).

Observe que la interpretación del estadístico “F” no dice nada acerca de si el error aleatorio del método evaluado es *aceptable*, sólo dice si es diferente al del método de comparación. El método comparativo en sí podría o no tener error aleatorio aceptable, haciendo así que medir la diferencia en el error aleatorio entre el método analítico evaluado y el de comparación no sea realmente relevante en al momento de evaluar la aceptabilidad. La aceptabilidad depende del tamaño del error aleatorio, independientemente de que sea menor o mayor que el error aleatorio del método de comparación.

**Tabla-F Valores Críticos de F para p = 0.05 (probabilidad) y
grados de libertad determinados (GL)**

GL del denominador	5	10	15	20	30	60	∞
1	230.00	242.00	246.00	248.00	250.00	252.00	254.00
2	19.30	19.40	19.40	19.40	19.50	19.50	19.50
3	9.01	8.79	8.70	8.66	8.62	8.57	8.53
4	6.26	5.96	5.86	5.80	5.75	5.69	5.63
5	5.05	4.74	4.62	4.56	4.50	4.43	4.36
6	4.39	4.06	3.94	3.87	3.81	3.74	3.67
7	3.97	3.64	3.51	3.44	3.38	3.30	3.23
8	3.69	3.35	3.22	3.15	3.08	3.01	2.93
9	3.48	3.14	3.01	2.94	2.86	2.79	2.71
10	3.33	2.98	2.85	2.77	2.70	2.62	2.54
15	2.90	2.54	2.40	2.33	2.25	2.16	2.07
20	25.71	2.35	2.20	2.12	2.04	1.95	1.84
30	2.53	2.16	2.01	1.93	1.84	1.74	1.62
60	2.37	1.99	1.84	1.75	1.65	1.53	1.39
∞	2.21	1.83	1.67	1.57	1.46	1.32	1.00

Análisis de Regresión Lineal

En el experimento de comparación de métodos, se recomienda que sean graficados primero los datos de la siguiente forma, los valores del método a ser evaluado en el eje de las ordenadas “Y” y los valores del método de comparación en el de las abscisas “X”. Puede dibujarse una línea imaginaria a través de los puntos generados a partir de los pares de datos, y por inspección visual evaluar cualitativamente los errores analíticos. Esta evaluación puede hacerse cuantitativamente usando un *análisis de regresión lineal*.

Este procedimiento estadístico es nombrado algunas veces como análisis de los cuadrados mínimos debido a la técnica utilizada para determinar la mejor localización de la línea recta a través de un conjunto de puntos de datos. Cuando se dibuja la línea manualmente, su locación se juzga minimizando visualmente las distancias de todos los puntos a la línea. Para encontrar la mejor línea posible, se pueden dibujar varias líneas y medirse las distancias de todos los puntos a cada una de las líneas. Las distancias del conjunto de puntos a cada línea podría ser sumada y la línea con el menor total debería ser la mejor línea. De hecho, la mejor línea es aquella que minimiza los cuadrados de las distancias de los puntos de datos a la línea, permitiendo que los puntos más distantes tengan un mayor efecto en la línea. Ésta línea se llama línea de los cuadrados mínimos y el proceso estadístico se llama análisis de los cuadrados mínimos.

La línea recta se describe por la *ecuación de regresión*:

$$Y_1 = a + bx_1$$

donde “a” es la intersección-Y (el punto donde la línea intersecta con el eje de las ordenadas “Y”) y “b” es la pendiente (el ángulo de la línea en el gráfico).

La *pendiente* (b) se calcula de la ecuación:

$$b = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2}$$

Note que el término $(\sum x_i)^2$ significa que los valores individuales son sumados y la suma se eleva al cuadrado. No es lo mismo que $\sum x_i^2$, que significa que los valores individuales se elevan al cuadrado primero para luego ser sumados.

El *intercepto en y* (a) se calcula de la ecuación:

$$a = \bar{y} - b \bar{x}$$

donde \bar{y} y \bar{x} son los valores promedio de las muestras de pacientes por el método analítico evaluado y el método de comparación respectivamente, y "b" es la pendiente de la ecuación de arriba.

La desviación estándar alrededor de la línea de regresión ($S_{x/y}$) es calculada a partir de la ecuación:

$$S_{y/x} = \sqrt{\frac{\sum(y_i - Y_i)^2}{N - 2}}$$

Esta estadística es más conocida como el *error estándar* y a veces como la *desviación estándar de los residuos*. Cuantifica la dispersión de los puntos alrededor de la línea de regresión. Para calcular $S_{y/x}$, se calculan primero los valores teóricos de Y correspondientes a los valores de x usando la ecuación de regresión ($Y_i = a + bx_i$). Las diferencias de los valores reales del examen son calculadas ($y_i - Y_i$), elevadas al cuadrado, y luego sumadas. Observe que la desviación estándar alrededor de la ecuación de la línea de regresión que vemos arriba es muy similar a la ecuación para calcular una desviación estándar. Hay dos diferencias. La primera es que Y_i es usada en lugar del valor de la media. Esto corresponde a tomar los puntos en la línea de regresión y calcular las diferencias o desviaciones de los puntos reales alrededor de la línea de regresión. Por esto es que preferimos llamar a la estadística desviación estándar *alrededor de la línea de regresión*. La segunda diferencia de la ecuación de la desviación estándar es el número de grados de libertad en el denominador, que aquí es $N-2$ en lugar de $N-1$. Esto ocurre debido a que tanto la pendiente como la intersección son calculadas antes de calcular $S_{x/y}$, así que hay dos restricciones en la serie de mediciones, dejando $N-2$ comparaciones independientes o grados de libertad sobrantes.

Otras dos estadísticas que algunas veces son útiles, aunque rara vez vistas en reportes de validación de métodos en la literatura, son la *desviación estándar de la pendiente* (S_b) y la *desviación estándar de la intersección* (S_a):

$$S_b = \frac{S_{y/x}}{\sqrt{\sum(x_i^2 - \bar{x})^2}}$$

$$S_a = S_{y/x} \sqrt{\frac{\sum x_i^2}{N \sum(x_i^2 - \bar{x})^2}}$$

Estas dos estadísticas cuantifican la dispersión de valores de la pendiente y la intersección, respectivamente, y permiten calcular los intervalos de confianza para la pendiente y la intersección.

Los valores ideales para las estadísticas de regresión serían 1.000 para la pendiente y cero para el resto. Estos valores ideales serían obtenidos sólo cuando los resultados del método analítico evaluado fueran exactamente los mismos que los del método de comparación. Las desviaciones de estas estadísticas de sus valores ideales son, por lo tanto, indicaciones de errores analíticos.

Una de las aplicaciones más útiles de las estadísticas de regresión es el cálculo del error sistemático. Esto puede ser realizado para cualquier concentración en el rango de concentraciones estudiado en el experimento de comparación de métodos. Por ejemplo, X_c podría ser un nivel de decisión médica, donde la interpretación de los resultados del examen es crítica. El valor del método analítico evaluado correspondiente, Y_c , podría ser calculado a partir de la ecuación de regresión, sustituyendo X_c por x_i e Y_c por Y_i en la ecuación, para dar:

$$Y_c = a + b X_c$$

El error sistemático (ES) está dado entonces por la ecuación:

$$ES = Y_c - X_c$$

cuando $X_c = \bar{X}$, Y_c será igual a \bar{Y} , y así el ES será igual al sesgo como se estimó a partir del estadístico “t”. Esto sugiere que un análisis de regresión proveerá la misma estimación del error sistemático, al menos en el caso especial donde se desea una estimación en la media de las concentraciones estudiadas.

Coeficiente de Correlación

El *coeficiente de correlación* (*r*) puede ser calculado a partir de la ecuación mostrada abajo, usando los términos de sumatoria de los cálculos de regresión:

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[N \sum x_i^2 - (\sum x_i)^2][N \sum y_i^2 - (\sum y_i)^2]}}$$

Los valores de “*r*” pueden variar de -1.000 a +1.000. Un valor de +1.000 indica una correlación perfecta, esto es, a medida que incrementa “x”, “y” incrementa proporcionalmente. Un valor de 0.000 indica que no hay correlación, esto es, a medida que incrementa “x”, “y” podría incrementar o disminuir.. Si “*r*” es -1.000, existe una correlación negativa perfecta. A medida que incrementa x, y disminuye proporcionalmente. No se encontrarán situaciones de correlación negativa en los experimentos de comparación de métodos, que es el experimento donde se ven usualmente aplicaciones de “*r*”.

9: ¿Cómo se determina el rango reportable de un método?

Este capítulo describe cómo puede ser determinado el rango reportable usando una serie de estándares con valores asignados o utilizando una serie de diluciones a partir de la muestra de un paciente o de una muestra obtenida a partir de un pool de pacientes. Se recomienda un análisis gráfico de los datos, junto con una valoración visual del rango de linealidad. Se puede utilizar la herramienta disponible *online* para graficar datos de linealidad para cualquiera de las opciones planteadas, estándares con valores asignados o la serie de diluciones a partir de una muestra o pool de muestras. Finalmente, el capítulo brinda una breve discusión del procedimiento experimental para la “verificación de la calibración”.

Objetivos:

- Aprender el procedimiento práctico para determinar el rango reportable.
- Identificar factores experimentales importantes.
- Calcular y graficar datos experimentales para determinar el rango reportable.
- Reconocer la diferencia entre el rango reportable y la verificación de la calibración.

Materiales del capítulo:

- VM — El experimento de Linealidad o Rango Reportable
- Problema: Datos de validación de un método para colesterol
- “Herramienta para graficar datos de linealidad” (www.westgard.com/mvtools.html)

Cosas qué hacer:

- Estudiar los materiales.
- Practicar usando las hojas de cálculo y las herramientas de cálculo disponibles *online*.
- Trabajar con el problema de colesterol.

Preguntas de Autoevaluación:

- ¿Cuántos niveles de materiales son necesarios para validar el rango reportable?
- ¿Cuál es una forma práctica de obtener una serie de esos materiales?
- ¿Cuántas repeticiones se efectúan por lo general sobre cada uno de los materiales?
- ¿Cómo se analizan los datos para evaluar el rango reportable?
- ¿Cómo se analizan los datos para evaluar linealidad?

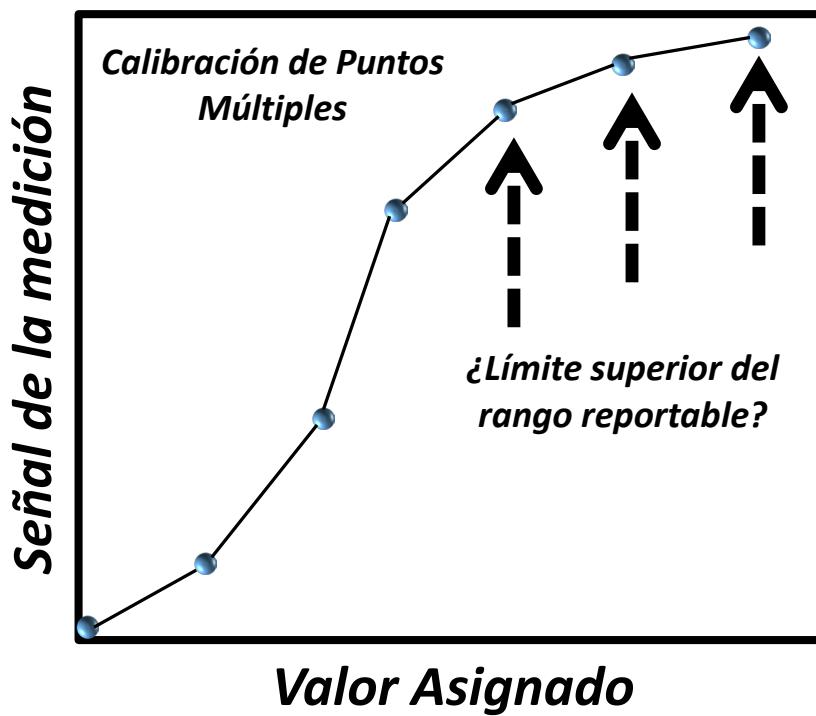
Validación de Métodos: El Experimento de Linealidad o Rango Reportable

James O. Westgard, PhD y Elsa Quam, BS, MT(ASCP)

Es importante determinar el rango reportable de un método de laboratorio, esto es, los resultados del examen más bajos y más altos que son confiables y pueden ser informados. Los fabricantes establecen especificaciones para el rango reportable asentando los límites inferior y superior del rango. Es muy importante verificar estas especificaciones declaradas por los fabricantes, particularmente cuando se asume que un método es lineal y se utiliza una “calibración de dos puntos”.

Los fabricantes proveen instrucciones y usualmente materiales para calibrar sus métodos. Los métodos podrían requerir de una calibración con cada corrida analítica, o sólo diaria, semanal o mensualmente, o incluso a intervalos mayores, o sólo cuando se ha puesto en uso un nuevo lote de reactivos, dependiendo de la documentación que dispone el fabricante sobre la estabilidad del método. Cuando se establece una calibración a largo plazo, podría ser necesario para el laboratorio verificar periódicamente que la calibración es correcta, esta es una función llamada “verificación de la calibración” en la Regla Final de CLIA.

La *calibración* es el procedimiento que determina la relación entre la señal generada por una metodología analítica y los resultados del examen que son informados. La “calibración de puntos múltiples” se requiere para métodos que no generan una respuesta lineal (por ej., métodos de inmunoensayo) y usualmente involucra analizar de tres a cinco (o incluso más) niveles de calibrador y utilizar una rutina de ajuste de curvas para establecer la función de calibración.



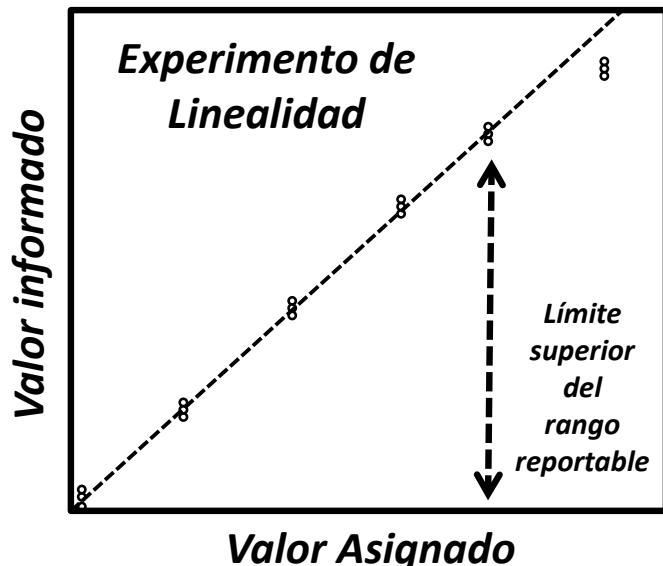
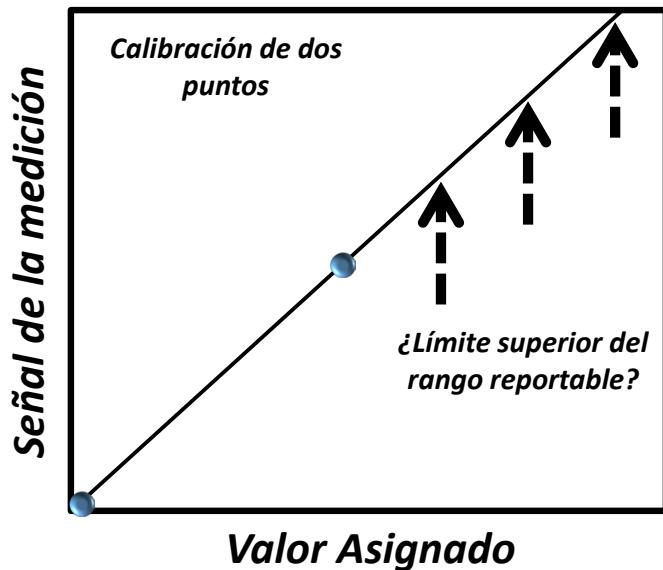
Podría ser requerido algo de criterio para evaluar el límite superior del rango reportable incluso cuando haya calibradores altos en el rango superior. El cambio en la respuesta podría volverse tan bajo en el límite superior que la resolución de la medición no sería ya satisfactoria.

Por ejemplo, ¿son significativos los cambios entre los resultados analíticos entre los dos calibradores más altos mostrados en la figura o debería restringirse el rango reportable al valor máximo del quinto calibrador?

En analizadores automáticos, se utiliza comúnmente una “calibración de dos puntos”. Típicamente, un calibrador provee un “punto cero” y el otro un “punto de referencia”, como se muestra aquí. Se asume que una función de calibración lineal puede ser dibujada entre el punto cero y el punto de referencia y que el rango lineal se extiende más allá de éste último. ¿Qué tan más allá? La pregunta debería ser contestada determinando el rango reportable.

El experimento a realizar es llamado a frecuentemente “experimento de linealidad”, aunque no hay técnicamente una condición que un método provea una respuesta lineal a menos que se utilice una calibración de dos puntos. CLIA utiliza el término “rango reportable” más que “rango lineal”, pero es común también referirse a este rango como el “rango lineal”, el “rango analítico” o el “rango de trabajo” para un método.

Debido a que la terminología varía en ésta área, es importante comprender los términos y los significados en las regulaciones de CLIA.



Definición de términos

La Regla Final de CLIA incluye las siguientes definiciones (493.2) que son importantes al momento de interpretar estos requisitos:

Calibración significa un proceso de análisis y ajuste de un instrumento o sistema analítico para establecer una correlación entre la respuesta de medición y la concentración o cantidad de la sustancia que está siendo medida por el procedimiento analítico.

Verificación de la calibración significa el análisis de materiales de concentración conocida en la misma manera que las muestras de pacientes para corroborar la calibración del instrumento o del sistema analítico a través del rango reportable para los resultados de los exámenes de los pacientes.

Note la relación al “rango reportable”, que también se define en la Regla Final de CLIA [1] como:

Rango reportable significa el rango de los valores de los resultados analíticos sobre los que el laboratorio puede establecer o verificar la exactitud del instrumento o de la respuesta de las mediciones del sistema analítico.

El *College of American Pathologists (CAP)* hace uso de términos diferentes [2]:

Rango analítico de medición (RAM) es el rango de los valores del analito que un método puede medir directamente en el espécimen sin ninguna dilución, concentración u otro tratamiento previo que no sea parte del proceso de ensayo usual.

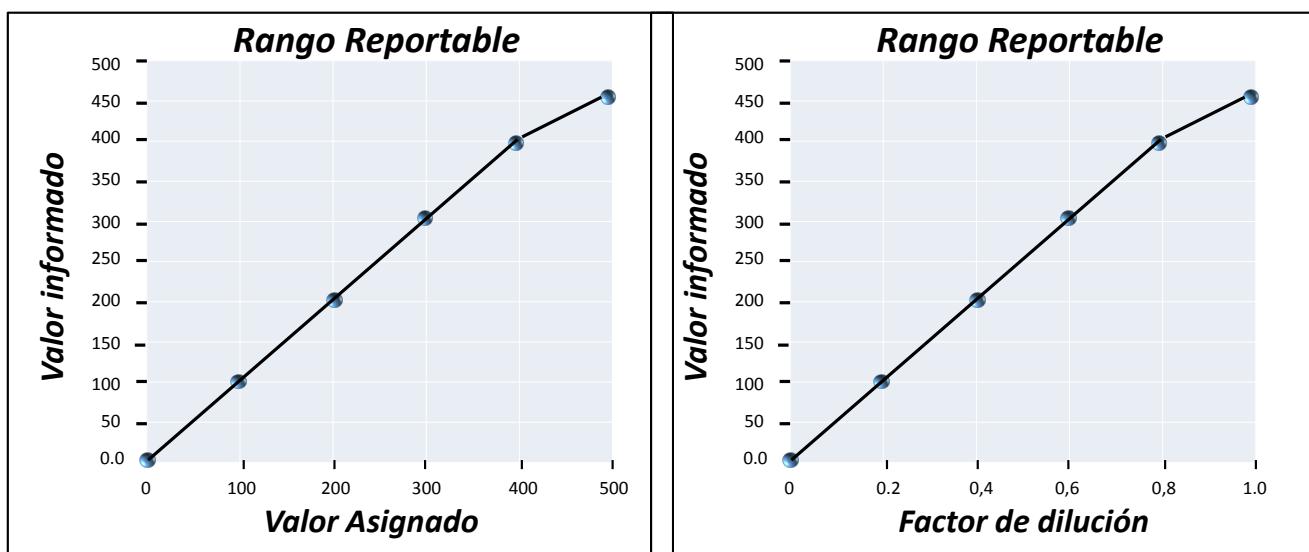
Rango clínico reportable (RCR) es el rango de los valores del analito que un método puede medir, permitiendo la dilución, la concentración u otro tratamiento previo del espécimen utilizado para extender el rango directo analítico de medición.

Reconocemos que RAM corresponde al rango reportable de CLIA y que RCR provee información práctica adicional que puede tomar en consideración el límite de detección de un método y/o la extensión del RAM por procedimientos verificados para diluciones de especímenes de pacientes. Por ejemplo, el límite inferior del RCR para un método como la Proteína C-Reactiva (PCR) podría estar restringido a reportar valores como menores a 10 mg/L, en lugar de cero, para evitar confusiones con resultados de un método de mayor sensibilidad (PCR ultrasensible) donde los valores de intereses están en un rango más bajo, de 1 a 5 mg/L. El límite superior del RCR podría permitir la dilución de especímenes del

paciente para extender los estimados de los valores analíticos más allá del límite superior del RAM, por ejemplo, para exámenes tales como glucosa y creatinina, donde podría emplearse un protocolo de dilución para estimar valores fuera del RAM.

Factores a considerar

El experimento de linealidad involucra una serie de muestras de concentraciones conocidas (valores asignados) o una serie de diluciones conocidas de un espécimen de muy alta concentración o un pool de pacientes. Los valores analíticos medidos o reportados son comparados a los valores asignados o a los valores de las diluciones, comúnmente graficando los valores medidos en el eje de las ordenadas “Y” y los valores asignados o de las diluciones en el de las abscisas “X”, como se muestra en la figura abajo.



El rango reportable se evalúa por inspección de la gráfica de linealidad. Esta inspección podría involucrar dibujar manualmente la mejor línea recta a través de la porción lineal de los puntos, dibujando una línea punto a punto a través de todos los puntos y luego comparándola con la mejor línea recta, o ajustando una línea de regresión a través de los puntos en el rango lineal. Hay cálculos estadísticos más complejos que se recomiendan en algunas guías, como el protocolo EP-6 de CLSI [3] para evaluar la linealidad de los métodos analíticos. Pero se acepta comúnmente que el rango reportable puede ser determinado adecuadamente de una valoración “visual”, esto es, dibujando manualmente la mejor línea recta que ajuste a los puntos más bajos de la serie.

Número de niveles

El Instituto de Estándares del Laboratorio Clínico (*CLSI*, por sus siglas en Inglés) recomienda un mínimo de 4 -preferiblemente 5- diferentes niveles de concentraciones [3]. Pueden ser usados más de 5, particularmente si el nivel superior del rango reportable necesita ser maximizado, pero hemos encontrado que 5 niveles son convenientes y casi siempre suficientes.

Materiales

Las soluciones estándares pueden ser fáciles de preparar para algunos exámenes. Para otros, los fabricantes y proveedores de esquemas de evaluación de la competencia podrían suplir un conjunto de soluciones para la evaluación de la linealidad con valores conocidos o relaciones conocidas entre materiales. También podrían usarse especímenes de pacientes o pools de especímenes de pacientes, que son a menudo convenientes y económicos cuando existen valores altos disponibles. En algunos casos, por ej., en Drogas de Monitoreo Terapéutico, podría ser necesario introducir en un pool el analito a ser medido para obtener el nivel alto deseado.

Diluyentes para utilizar cuando se trabaja con especímenes de pacientes

La selección del diluyente es importante cuando se necesita mantener la matriz del espécimen, por ej., cuando el suero es el tipo usual de espécimen, podría ser importante mantener una matriz sérica para la serie de muestras analíticas diluidas. Para exámenes generales de química clínica, es frecuente utilizar agua o solución salina como diluyente. Para otras pruebas, podría ser mejor utilizar preparaciones de albúmina bovina o sérica, especímenes con bajas concentraciones o suero libre de fármacos. Una forma de escoger el diluyente es seguir la recomendación del fabricante para diluir especímenes de pacientes fuera de rango.

Procedimiento para preparar diluciones

Es conveniente usar dos pools — uno cerca del nivel cero o cerca del límite de detección y el otro cerca o ligeramente por encima del límite superior esperado del rango reportable. Determine el volumen total necesario para los análisis, seleccione pipetas volumétricas apropiadas y luego haga lo siguiente:

1. Etiquete el pool de concentración más baja como “Pool 1”, y el de más alta concentración como “Pool 5”.

2. Prepare la Mezcla 2 (75/25) con 3 partes del Pool 1 + 1 parte del Pool 5.
3. Prepara la Mezcla 3 (50/50) con 2 partes del Pool 1 + 2 partes del Pool 5.
4. Prepare la Mezcla 4 (15/75) con 1 parte del Pool 1 + 3 partes del Pool 5.

Si se desean más niveles, éste protocolo de dilución puede ser modificado, por ej., los dos *pools* pueden ser mezclados 4 a 1, 3 a 2, 2 a 3 y 1 a 4 para dar cuatro niveles intermedios para un total de seis niveles para el experimento.

Número de replicados de medición

El CLSI recomienda hacer 4 mediciones de cada espécimen o pool. En la práctica, hemos encontrado que 3 replicados son generalmente suficientes, lo que implica procesar por triplicados los pools originales alto y bajo también.

Análisis de datos

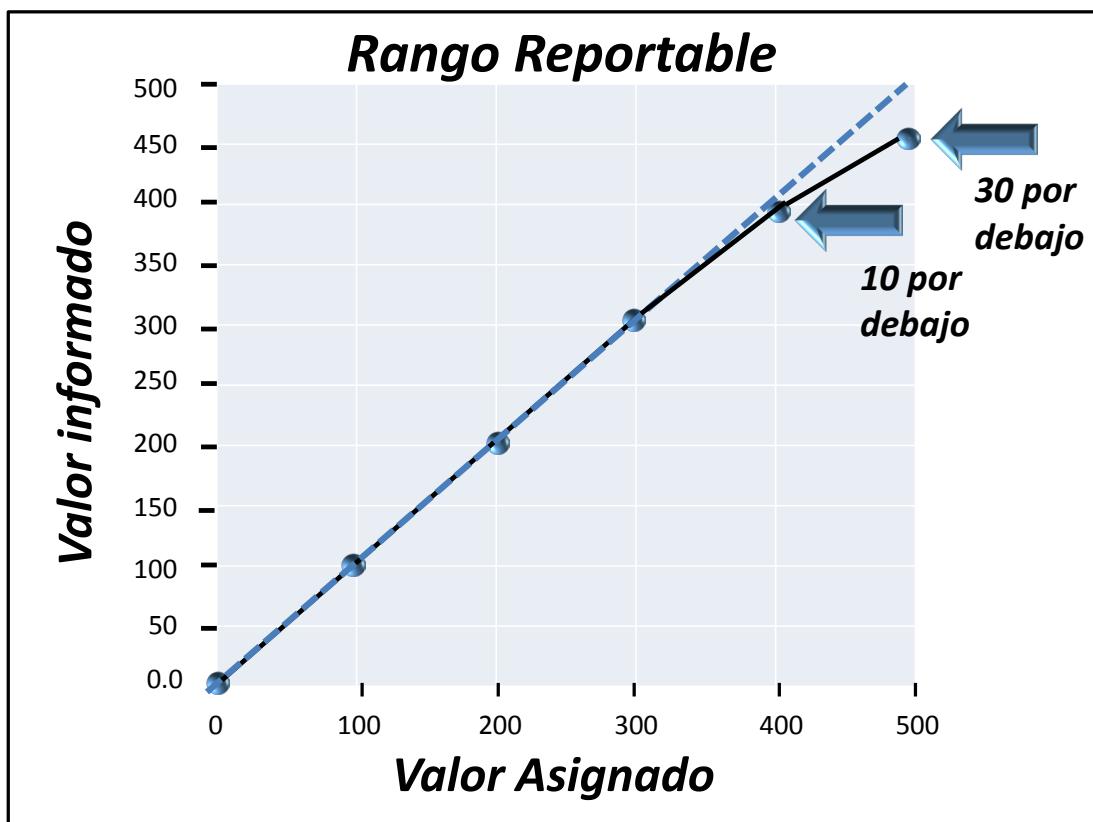
Grafique la media de los valores medidos en el eje de las ordenadas “Y” contra los valores asignados o valores relativos o factores de dilución en el eje de las abscisas “X”. Primero dibuje una línea punto a punto a través del rango analítico entero. Luego dibuje manualmente la mejor línea recta a través de la mayor cantidad posible de puntos, asegurándose de que la línea pasa por los puntos más bajos o los estándares más bajos o los valores de dilución más bajos. En las concentraciones donde la línea recta ya no pasa por los puntos, se estima el error sistemático debido a la no linealidad. Compare ese error sistemático más el error aleatorio esperado en esa concentración (2 S) con el error total permitido para el examen.

Ejemplo de colesterol

Los datos son los siguientes:

Asignado: 0, observado: 0, 5, 10, promedio: 5.0;
Asignado: 100, observado: 95, 100, 105, promedio: 100;
Asignado: 200, observado: 200, 195, 205, promedio: 200;
Asignado: 300, observado: 310, 300, 290, promedio: 300;
Asignado: 400, observado: 380, 390, 400, promedio: 390;
Asignado: 500, observado 470, 460, 480, promedio: 470.

La figura abajo muestra los valores promedio (valores observados) graficados en el eje de las ordenadas “Y” contra los valores asignados en el eje de las abscisas “X”.



La línea sólida representa la línea dibujada punto a punto y la línea punteada representa la línea recta ajustada a los puntos en el área baja a media del rango. Se estiman diferencias sistemáticas de 0 mg/dL en 300 mg/dL, 10 mg/dL en 400 mg/dL, y 30 mg/dL en 500 mg/dL. El rango reportable se extiende claramente a 300 mg/dL, pero ¿se extiende a 400 mg/dL o 500 mg/dL?

- En 500 mg/dL, dado un método con un CV de 3.0%, la desviación estándar sería 15 mg/dL y el estimado del error aleatorio a 2 S sería 30 mg/dL. Esto significa que en una muestra con un valor verdadero de 500, en promedio sería observado un valor de 470 mg/dL debido al error sistemático producto de la no linealidad. Además, ese valor podría ser + 30 mg/dL debido al error aleatorio, esto es, el valor esperado estaría en el rango de 440 a 500 mg/dL para una muestra con un valor verdadero de 500 mg/dL. Dado que el criterio de CLIA para el error total permitido es del 10%, que es 50 mg/dL a un nivel de 500 mg/dL, los errores que serían observados a 500 mg/dL podrían ser mayores que el error permitido, así que el rango reportable debería ser restringido a una concentración menor.

- A 400 mg/dL, la desviación estándar sería de 12 mg/dL, dando un estimado del error aleatorio a 2 S es igual a 24 mg/dL. En una muestra con un valor verdadero de 400 mg/dL, en promedio sería observado un valor de 390 mg/dL debido al error sistemático producto del error de no linealidad. Una adición del error aleatorio da un rango esperado de 366 a 414 mg/dL, que significa que un resultado podría tener un error de hasta 34 mg/dL. El criterio de CLIA del 10% provee un error total permitido de 40 mg/dL a 400 mg/dL, así que los resultados esperados son correctos considerando el error total permitido que se a escogido para el método ($34 \text{ mg/dL} < 40 \text{ mg/dL}$), y por tanto el rango reportable se extiende a 400 mg/dL.

Rango Reportable y Verificación de la Calibración

De acuerdo con CLIA, la verificación de la calibración implica examinar materiales de concentración conocida en la misma manera que se examina la muestra de un paciente para corroborar la calibración del instrumento o del sistema analítico a través del rango reportable para los resultados de los exámenes de pacientes. Para sistemas que son calibrados de forma infrecuente, la calibración debe ser verificada al menos cada 6 meses y más asiduamente si hay un cambio completo de reactivos, un cambio de partes o componentes críticos, un mantenimiento mayor o un problema con el control estadístico interno de la calidad incorregible que demuestre una modificación en los resultados analíticos.

Si bien el experimento de verificación de la calibración se parece mucho al experimento para el rango reportable, hay una diferencia importante — el experimento debe usar materiales con valores asignados. Los materiales podrían ser una serie especial calibradores de fabricantes o vendedores de reactivos, calibradores que se corren como desconocidos, es decir como si fuesen muestras de pacientes,, muestras control con valores asignados, muestras que participan en esquemas de evaluación de la competencia con valores conocidos, e incluso muestras de pacientes con valores ensayados [4]. La disposición de CLIA para examinar materiales como muestras de pacientes no debería considerarse una restricción del diseño experimental. El uso de mediciones replicadas brindará mejor información acerca de errores sistemáticos posibles originados por la calibración.

El procedimiento de análisis de datos debería determinar las diferencias entre los valores observados y asignados, luego comparar aquellas diferencias con un criterio para la aceptabilidad, previamente definido por el laboratorio. El enfoque puede ser similar al esbozado aquí para el rango reportable si se hacen los replicados de las mediciones y se determinan los valores de la media de esos valores. Si se realizan mediciones únicas, entonces las diferencias observadas podrían ser comparadas directamente al error total permitido definido por CLIA.

Referencias

1. US Centers for Medicare & Medicaid Services (CMS). Medicare, Medicaid, and CLIA Programs: Laboratory Requirements Relating to quality Systems and Certain Personnel Qualifications. Final Rule. Fed Regist Jan 24 2003;16:3640-3714.
2. CAP Laboratory Accreditation Checklists. <http://www.cap.org>
3. CLSI EP6. Linearity of the Linearity of Quantitative Measurement Procedures: A Statistical Approach. Clinical Laboratory Standards Institute, Wayne PA, 2003.
4. CMS State Operations Manual – Interpretative Guidelines: Appendix C. http://www.cms.hhs.gov/CLIA/03_Interpretive_Guidelines_for_Laboratories.asp

Problema- Datos de la validación de un método para colesterol

Una serie de soluciones estándar con valores asignados de 50, 100, 150, 200, 250, 300, 350, y 400 mg/dL fueron medidas por triplicado. Prepare un gráfico a partir de los datos: evalúe el rango reportable de este método. (Las respuesta aparecen en la página 306)

Valores Asignados	Valores Obtenidos		
	Resultado 1	Resultado 2	Resultado 3
50	50	50	49
100	102	101	101
150	148	159	147
200	190	200	198
250	237	261	251
300	300	314	310
350	345	352	343
400	396	395	372

10: ¿Cómo se determina la imprecisión de un método?

Una vez que conoce el rango analítico para informar resultados de pacientes, es importante determinar la imprecisión del método en concentraciones donde es crítica la interpretación de los resultados de los pacientes. Este capítulo describe cómo realizar un experimento de replicación, los factores que son importantes en el procedimiento experimental y los cálculos y gráficos que son apropiados para los datos.

Objetivos:

- Aprender un procedimiento práctico para estimar la imprecisión de un método.
- Identificar los factores experimentales importantes.
- Calcular los datos experimentales para valorar la imprecisión.

Materiales del capítulo:

- VM — El experimento de replicación, por James O. Westgard, PhD
- Problema: Datos de validación de un método de colesterol

Cosas qué hacer:

- Estudiar los materiales.
- Practicar usando la herramienta de cálculo *online*.
- Trabajar con el problema de colesterol.

Preguntas de Autoevaluación:

- ¿Cuántos niveles o materiales se necesitan para un experimento de replicación?
- ¿Cuál es el número mínimo de mediciones necesarias para cada material o nivel?
- ¿Qué período de tiempo debería abarcar el experimento de replicación final?
- ¿Cómo se analizan los datos para estimar la imprecisión?

Validación de Métodos: Experimento de Replicación

James O. Westgard, PhD

Se realiza un experimento de replicación para estimar la imprecisión o error aleatorio del método analítico. Los métodos de medición casi siempre están sujetos a variaciones aleatorias. Recuerde nuestro ejemplo de la báscula de baño de errores aleatorios y sistemáticos en *El plan experimental*, así como las descripciones gráficas en *El significado secreto, interno, oculto y más profundo*. Repetir las mediciones usualmente revelará resultados ligeramente diferentes, algunas veces un poco mayores, algunas veces un poco menores. Determinar la cantidad de error aleatorio es usualmente uno de los primeros pasos en un estudio de validación de métodos.

Un experimento de replicación se realiza típicamente al obtener los resultados de 20 muestras del mismo material y luego calculando la media, la desviación estándar y el coeficiente de variación. El propósito es observar la variación esperada en un resultado de examen bajo las condiciones normales de operación del laboratorio. Idealmente, la variación del examen debería ser pequeña, esto es, todas las respuestas en las mediciones repetidas deberían arrojar resultados similares.

El experimento de replicación estima el error aleatorio causado por factores que varían en la operación del método, tales como el pipeteo de las muestras, las condiciones de reacción que dependen del tiempo, el mezclado, la temperatura y el calentamiento, e incluso las mediciones en sí mismas. En sistemas no automatizados, la variación en las técnicas aplicadas por distintos analistas podría contribuir mucho a la variación observada de un examen. Con sistemas automatizados, la falta de uniformidad y la inestabilidad del instrumento y de las condiciones de reacción podrían causar pequeñas variaciones que podrían verse nuevamente como variaciones positivas y negativas en los resultados finales del examen. Mientras que el efecto exacto no puede ser predicho en cualquier momento, la distribución de estos efectos a lo largo del tiempo puede ser predicha para describir qué tan grande podría ser el error aleatorio.

Factores a considerar

La cantidad de error aleatorio que será observada depende del diseño experimental debido a que ciertas variables podrían no manifestarse a menos que se escojan las condiciones apropiadas. Por ejemplo, cuando se realiza un experimento en un periodo corto de tiempo, digamos dentro de una misma corrida analítica, los

efectos de variación a largo plazo debidos a los cambios día a día en las condiciones de operación no serán observados. La temperatura ambiente podría permanecer constante en ese periodo corto de tiempo y variar más a lo largo de un día, en diferentes días y en diferentes estaciones. Son factores importantes para diseñar el experimento: el periodo de tiempo del experimento, la matriz de las muestras a ser analizadas, el número y concentración de materiales a ser analizados y el número de muestras a ser analizadas. Mientras que se espera que el número de analistas que realicen el examen también sea un factor, esta variable se controla generalmente durante los estudios de validación de métodos y sólo uno o pocos analistas bien entrenados están involucrados en estos estudios.

Periodo de tiempo del experimento

La longitud de tiempo en la que es conducido el experimento es crítica para la interpretación de los datos y la conclusión que podría ser obtenida a partir de los mismos. Cuando se analizan muestras en una sola corrida analítica, el error aleatorio observado “intracorrida” será generalmente menor (y optimista) debido a que los resultados sólo son afectados por aquellos factores que varían en este periodo de tiempo corto. Estas condiciones llevarían al método a obtener el mejor desempeño posible; si este desempeño no es aceptable, el método debería ser rechazado o es necesario identificar las causas que generan estos errores aleatorios y eliminarlas antes de que se lleven a cabo más exámenes.

Un experimento llevado a cabo en un periodo de tiempo de un día, esto es, “intradía”, usualmente mostrará más variación que un experimento “intracorrida” a menos que el método esté altamente automatizado y sea muy estable. Se espera que un experimento conducido en un periodo de tiempo de veinte días provea un estimado incluso más realista de la variación que será observada en las muestras de los pacientes a lo largo del tiempo. Este estimado podría ser llamado imprecisión “día a día”, “entre días” o “total” del método. La CLSI prefiere el término “imprecisión total” tal vez porque implica que están incluidos los componentes de la variabilidad “intradía” y “entre días” [1].

Matriz de la muestra

Los otros materiales presentes en una muestra constituyen su matriz. Por ejemplo, la matriz de interés para un examen de laboratorio podría ser sangre entera, suero, orina o líquido cefalorraquídeo. Mientras que podría ser de interés medir glucosa en cada uno de estos tipos de especímenes, será difícil encontrar un único método para todos ellos. Al evaluar el desempeño de un método, es importante usar muestras que tengan una matriz tan cercana como sea posible al tipo real del espécimen de interés para el método evaluado.

Las muestras para examen están comúnmente disponibles como soluciones estándar, soluciones control, *pools* de pacientes, y muestras individuales de pacientes. Todas pueden ser usadas en un experimento de replicación, pero cada una tiene ciertas ventajas al igual que limitaciones.

Las *soluciones estándar* están disponibles para analitos comunes en química clínica y pueden prepararse para lograr las concentraciones de interés. La matriz de las soluciones estándar es usualmente más simple que la de las muestras reales de pacientes, por ejemplo, el estándar podría ser acuoso y la muestra del paciente ser suero con una alta concentración proteica. Así, un estimado del error aleatorio en una solución estándar sería optimista y es más probable que represente el mejor desempeño posible. Aun así, si este desempeño obtenido en condiciones tan favorables no es satisfactorio, se puede tomar la decisión de rechazar el método.

Las *soluciones control o materiales de control* pueden ser obtenidos de fuentes comerciales en forma y tamaño convenientes y con estabilidad a largo plazo. La matriz podría ser muy similar a la matriz de las muestras de pacientes, pero podrían presentarse ciertos efectos debidos al agregado de estabilizantes, al efecto provocado por la liofilización y reconstitución, y el agregado de aditivos especiales para incrementar los niveles de ciertos analitos, tales como enzimas y lípidos. Podría volverse más difícil en el futuro obtener materiales de control preparados a partir de muestras de pacientes reales debido a la necesidad de examinar y documentar que estén libres de enfermedades infecciosas. Ver *Prácticas Básicas de Control de la Calidad* para una discusión más extensa sobre materiales de control.

Los pools de muestras frescas de pacientes se pueden usar a menudo para exámenes a corto plazo, particularmente en estudios de replicación “intracorrida” e “intradía”. Se pueden analizar diariamente duplicados de muestras frescas de pacientes por largos periodos de tiempo, pero estas muestras aún reflejarán sólo los componentes de la imprecisión “intracorrida” e “intradía”. El componente “entre días” no será observado a menos que se realicen duplicados en diferentes días, en cuyo caso se debe demostrar la estabilidad de la muestra fresca para el periodo de tiempo transcurrido entre los duplicados.

Número de materiales y concentraciones a ser analizadas

El número de materiales a ser analizados debería depender de las concentraciones que son críticas para el uso clínico del examen. Generalmente, dos o tres materiales deberían ser seleccionados para tener concentraciones de analitos a niveles importantes de decisión médica. Un nivel de decisión médica representa

una concentración donde la interpretación médica de un resultado de examen es crítica.

Para colesterol, los niveles de decisión médica son 200 mg/dL y 240 mg/dL de acuerdo a las recomendaciones del NCEP para interpretar un resultado de un examen de colesterol [2]. La glucosa tiene una interpretación crítica a varios niveles de decisión médica, tales como 50 mg/dL para hipoglucemia, 120 mg/dL para una muestra en ayuno, 160 mg/dL para una curva de tolerancia, y a niveles mayores como 300 mg/dL para monitorear pacientes diabéticos. Para lineamientos e información sobre niveles de decisión médica para una gran variedad de exámenes vea las recomendaciones para Niveles de Decisión Médica provistas por el Dr. Bernard Statland [3].

Número de muestras a evaluar

Es comúnmente aceptado que un mínimo de 20 muestras debería ser medido en el periodo de tiempo de interés. Un gran número de muestras dará un mejor estimado del error aleatorio, pero las consideraciones de costos y tiempo a menudo indican que los datos se evalúen en un periodo de tiempo temprano o más corto, luego se podrá juntar más datos si es necesario.

Cálculo de los datos

El error aleatorio se describe cuantitativamente calculando la media (\bar{X}), la desviación estándar (S) y el coeficiente de variación (CV) del número “n” de mediciones individuales “ $\sum X_i$ ”, usando estas ecuaciones:

$$\bar{X} = \frac{\sum X_i}{n}$$

$$S = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

$$CV = \left(\frac{S}{\bar{X}} \right) 100$$

Los programas de cálculo están disponibles en calculadoras, hojas de cálculo, programas estadísticos y software especializado de validación de métodos, así como en las herramientas de cálculo *online* disponibles en la página web de Westgard: <http://www.westgard.com/mvtools.html>. También es útil preparar un

histograma de los resultados para mostrar visualmente la variación aleatoria esperada y demostrar qué tan grande podría llegar a ser para una medición individual.

Para especímenes de pacientes analizados por duplicado, la desviación estándar se calcula a partir de las diferencias “d” entre duplicados, utilizando la siguiente ecuación, donde “n” es el número de diferentes especímenes de pacientes:

$$S_{\text{dup}} = \sqrt{\frac{\sum d^2}{2n}}$$

Criterios para un desempeño aceptable

Aunque los cálculos de estos datos son simples, evaluar si la desviación estándar calculada representa un desempeño analítico aceptable no es tan simple. El criterio sobre aceptabilidad depende de qué cantidad de error analítico es permitido sin afectar o limitar el uso e interpretación de un resultado de examen [4,5]. Como un punto de inicio para definir la cantidad de error permitido, recomendamos usar los criterios CLIA para aceptabilidad. Estos están disponibles en el Apéndice 1.

Para imprecisión a corto plazo, la desviación estándar “intracorrida” ($S_{\text{intra-c}}$) o la desviación estándar “intradía” ($S_{\text{intra-d}}$) debería ser 1/4 o menos del error total permitido (TE_a) definido como aceptable, esto es:

$$S_{\text{intra-c}} \text{ o } S_{\text{intra-d}} < 0.25 TE_a$$

Para imprecisión a largo plazo, la desviación estándar total (S_{tot}) debería ser 1/3 o menos del error total permitido (TE_a) definido, esto es:

$$S_{\text{tot}} < 0.33 TE_a$$

Se pueden hacer juicios semejantes sobre aceptabilidad usando un Cuadro de Decisión del Método, disponible en formato PDF y Excel en la página web de Westgard, esta alternativa será discutida en el capítulo 16.

Verificación del Reporte del Fabricante

CLIA no requiere estrictamente que un laboratorio determine la aceptabilidad del desempeño. En su lugar, CLIA dice que el laboratorio debería verificar las especificaciones provistas por los fabricantes (en el inserto por lo general) para precisión. Esto puede realizarse con el estadístico “F”, como se muestra a continuación:

- Obtenga la S esperada y el número de mediciones utilizado en el experimento de replicación aplicado por al fabricante al momento de declarar sus especificaciones (usualmente estos datos están incluidos en la documentación del instrumento o insertos de los reactivos), por ejemplo, $S = 3 \text{ mg/dL}$ basada en 31 mediciones.
- Obtenga la S y el número de mediciones de su experimento de replicación, por ejemplo, $S = 4 \text{ mg/dL}$ basada en 21 mediciones.
- Calcule el valor de “ F ”, la S mayor al cuadrado dividida por la S menor al cuadrado, esto es, $(4)^2/(3)^2 = 16/9 = 1.78$.
- Busque el valor crítico de “ F ” para 20 grados de libertad ($L = N - 1$) en el numerador y $L = 30$ en el denominador en la tabla “ F ” en la página 99, donde el valor encontrado debería ser 1.93.
- En este caso, la “ F ” calculada es menor que el valor crítico de “ F ”, lo que indica que no existe diferencia real entre la S observada en el laboratorio y la S reportada por el fabricante.
- Conclusión — ¡las especificaciones de desempeño del fabricante para precisión han sido verificadas!

Estudios mínimos recomendados

Seleccione al menos 2 materiales de control diferentes que representen las concentraciones de decisión médica baja y alta del examen de interés. Analice 20 muestras de cada material dentro de una corrida o dentro de un día para obtener un estimado a corto plazo de la imprecisión. Calcule la media, la desviación estándar y el coeficiente de variación para cada material. Determine si la imprecisión a corto plazo es aceptable antes de proceder con más exámenes.

Analice 1 muestra de cada uno de los materiales en 20 días diferentes para estimar la imprecisión a largo plazo. Calcule la media, la desviación estándar y el coeficiente de variación para cada material. Determine si la imprecisión a largo plazo es aceptable.

Consideraciones futuras

Podrían emplearse diseños experimentales algo más elaborados para proveer información más extensa sobre los componentes a corto y a largo plazo de la variación. Estos diseños a menudo usan cálculos estadísticos conocidos como Análisis de Varianza (ANOVA), como se ilustra en los protocolos de CLSI para precisión [1,7]. El protocolo EP5 está diseñando primariamente para fabricantes, mientras que el protocolo EP15 está diseñado para usuarios de laboratorio. El EP15 especifica un mínimo de 3 replicados por día por un periodo de 5 días,

dando un total de 15 mediciones. Para estimar la imprecisión total, los cálculos proveen un modo de determinar los componentes de variación “intracorrida” e “intercorrida” que luego se combinan para dar la imprecisión “total”. Los cálculos necesarios se presentan en una hoja de trabajo contenida en un apéndice del EP15, además hay un ejemplo a partir de un conjunto de datos que muestra cómo verificar las especificaciones de desempeño provistas por el fabricante. Dado que el EP15 permite la verificación de la precisión con datos de sólo 5 días, es probable que éste se convierta en el protocolo predeterminado en muchos laboratorios. Por ésta razón, se discute el EP15 a detalle en el Capítulo 19.

Referencias

1. CLSI CP5-A. Evaluation of Precision Performance of Quantitative Measurement Methods. Clinical and Laboratory Standards Institute, Wayne, PA 2004.
2. National Cholesterol Education Program Laboratory Standardization Panel. Current status of blood cholesterol measurement in clinical laboratories in the United States. Clin Chem 1988;34:193-201.
3. Statland BE. Clinical Decision Levels for Laboratory Tests, 2nd ed. Oradell NJ;Medical Economics Books, 1987.
4. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974;20:825- 833.
5. Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. Clin Chem 1990;36:1629-1632.
6. Westgard JO. A method evaluation decision chart (MEDx Chart) for judging method performance. Clin Lab Science 1995;8:277-283.
7. CLSI EP15-A2. User Verification of Performance for Precision and Trueness. Clinical Laboratory Standards Institute, Wayne, PA, 2006.

Problema — Datos de Validación de un Método de Colesterol

Replicación (usada para estimar la imprecisión o el error aleatorio)

Se corrieron dos materiales de control diferentes una vez al día por un periodo de 20 días. Calcule la media, desviación estándar (S) y el coeficiente de variación (%CV) para cada material de control. Prepare un histograma para cada control. (Las respuestas aparecen en la página 306).

Día	Control A	Control B
1	203	240
2	202	250
3	204	235
4	201	248
5	197	236
6	200	234
7	198	242
8	196	244
9	206	243
10	198	240
11	196	233
12	192	237
13	205	243
14	190	234
15	207	241
16	198	240
17	201	249
18	195	232
19	209	231
20	186	241

11: ¿Cómo se determina la inexactitud o sesgo de un método?

Este capítulo describe cómo se realiza un experimento de “comparación de métodos” y cómo deben ser analizados los datos, enfatizando en los cálculos estadísticos y los gráficos. Se consideran estadísticas de regresión y el estadístico “t”, junto con gráficos de comparación y diferencias.

Objetivos:

- Aprender un procedimiento práctico para estimar la inexactitud o el sesgo de un método.
- Identificar los factores experimentales importantes.
- Comparar análisis estadísticos y presentaciones gráficas que son apropiados para conjuntos de datos que cubran rangos analíticos estrechos y amplios.

Materiales del capítulo:

- VM — El experimento de comparación de métodos,
por James O. Westgard, PhD
- Problema: Datos de validación de un método de colesterol

Cosas qué hacer:

- Estudiar los materiales.
- Practicar usando la herramienta de cálculo *online*.
- Trabajar con el problema de colesterol.

Preguntas de Autoevaluación:

- ¿Cuál es el número mínimo de especímenes que se utilizan generalmente un estudio de comparación de métodos?
- ¿Cuántos replicados de las mediciones se efectúan por lo general?
- ¿Qué período de tiempo se recomienda para la duración del experimento?
- ¿Cómo deberían ser calculados los datos cuando existen dos niveles de decisión médica?
- ¿Cómo deben graficarse los datos si hay dos o más niveles de decisión médica?
- ¿Qué cálculos deben ser utilizados si hay un único nivel de decisión médica?
- ¿Cómo se grafican los datos cuando existe un único nivel de decisión médica?

Validación de Métodos: El Experimento de Comparación de Métodos

James O. Westgard, PhD

Un experimento de comparación de métodos se realiza para estimar la inexactitud o el error sistemático. Revise *El plan experimental* para ver cómo se ajusta este experimento con los otros. Se realiza este experimento analizando muestras de pacientes con el nuevo método (método analítico a evaluar) y un método de comparación, luego se estiman los errores sistemáticos en base a las diferencias observadas entre los métodos. Las diferencias sistemáticas en concentraciones críticas, que representan niveles de decisión médica, son los errores de interés. Sin embargo, la información sobre la naturaleza constante o proporcional del error sistemático también es útil y a menudo está disponible a partir de los cálculos estadísticos apropiados. Tanto el diseño experimental como los cálculos estadísticos son críticos para obtener estimados confiables de los errores sistemáticos.

Factores a Considerar

Método comparativo

El método analítico utilizado para la comparación debe ser cuidadosamente seleccionado debido a que la interpretación de los resultados experimentales dependerá de la asunción que pueda hacerse acerca de la exactitud de los resultados del método comparativo. Cuando sea posible, se debería escoger un “método de referencia” como método comparativo. Este término tiene un significado específico que infiere un método de alta calidad cuyos resultados se sabe son correctos a través de estudios comparativos con un “método definitivo” exacto y/o a través de su trazabilidad con materiales estándar de referencia. Cualquier diferencia entre un método analítico en evaluación y un método de referencia se le asigna al método analítico en evaluación, esto es, los errores se atribuyen al método analítico evaluado porque la veracidad del método de referencia está bien documentada.

El término “método comparativo” es un término más general y no implica que la veracidad del método haya sido documentada. La mayoría de los métodos de rutina del laboratorio caen en esta categoría. Cualquier diferencia entre un método analítico a evaluar y un método de rutina debe ser interpretada cuidadosamente. Si las diferencias son pequeñas, entonces los dos métodos tienen

la misma exactitud relativa. Si las diferencias son grandes e inaceptables desde un punto de vista clínico, entonces es necesario identificar cuál método es inexacto. Se pueden emplear experimentos de recuperación e interferencia para obtener información adicional.

Número de especímenes de pacientes

Un mínimo de 40 especímenes de pacientes diferentes deberían ser examinados por los dos métodos [1]. Estos especímenes deberían ser seleccionados para cubrir el rango reportable entero del método y deberían representar el espectro de enfermedades que se esperan en la aplicación de rutina del método. El *número* real de especímenes examinados es menos importante que la *calidad* de esos especímenes. Veinte especímenes cuidadosamente seleccionados en base a sus concentraciones probablemente proveerán mejor información que cien especímenes recibidos aleatoriamente en el laboratorio. La calidad del experimento y las estimaciones de los errores sistemáticos dependerán más de elegir un amplio rango de resultados de examen (calidad de los especímenes) que que de trabajar con un gran número de resultados.

La ventaja principal de realizar el experimento con un gran número de especímenes es identificar muestras de pacientes cuyos resultados no concuerden debido a interferencias en la matriz de esa única muestra. Esto es a menudo de interés cuando el nuevo método hace uso de una reacción química diferente o de un principio de medición diferente. Se recomiendan trabajar con un gran número de especímenes — de 100 a 200 — para evaluar si la especificidad del nuevo método es similar a la del método comparativo.

Mediciones únicas contra mediciones por duplicado

La práctica más frecuente consiste en analizar cada espécimen una vez por el método analítico evaluado y una vez por el método comparativo. Sin embargo, existen ventajas en hacer mediciones por duplicado cada vez que sea posible. Idealmente, estos duplicados deberían ser dos muestras diferentes (o tubos) que se analizan en diferentes corridas, o al menos en diferente orden (en lugar de replicados efectuados uno tras otro en el mismo tubo o copa de muestra). Los duplicados brindan una comprobación de la validez de las mediciones por los métodos individuales y ayudan a identificar problemas que surgen de mezclar muestras, errores de transposición y otra tipo de equivocaciones. Una o dos equivocaciones de este tipo pueden tener un fuerte impacto en las conclusiones obtenidas a partir del experimento. Este tipo de errores o equivocaciones causan mucha confusión al tener que definir si estos resultados discrepantes representan el desempeño del método, o si son “valores atípicos” que deberían ser retirados del conjunto de datos. Los análisis por duplicado pueden evidenciar si estas discrepancias observadas son consistentes o no.

Si no se realizan duplicados, entonces es de suma importancia inspeccionar los resultados de comparación en el momento en que son obtenidos, identificar aquellos especímenes donde las diferencias son grandes, y repetir los análisis mientras aún están disponibles los especímenes.

Periodo de tiempo

Deberían incluirse varias corridas analíticas diferentes en días diferentes para minimizar cualquier error sistemático que podría ocurrir en una sola corrida. Se recomienda un mínimo de 5 días [1], pero sería preferible extender el experimento a un periodo de tiempo más largo. Dado que el estudio de replicación a largo plazo probablemente se extenderá por 20 días, el estudio de comparación puede cubrir un periodo de tiempo similar y requerirá de sólo 2 a 5 especímenes de pacientes al día.

Estabilidad del espécimen

No más de dos horas deberían transcurrir entre que se analizan los especímenes en uno y otro método analítico (método analítico evaluado y método de comparación), a menos que se sepa que los especímenes tienen corta estabilidad, por ejemplo, el amonio y el lactato. Se podría mejorar la estabilidad para algunos exámenes añadiendo conservantes, separando el suero o el plasma de las células, refrigerando o congelando. El manejo de los especímenes debe ser definido cuidadosamente y sistematizado antes de comenzar el estudio de comparación de métodos. De otro modo, las diferencias observadas podrían ser debidas a variables en el manejo de especímenes, más que a los errores analíticos sistemáticos que son el propósito del experimento.

Análisis de datos

¡Aquí es dónde se pone difícil! Existe mucho debate y discusión sobre la forma correcta de analizar los datos obtenidos a partir de un experimento de comparación de métodos [2]. Esto ha sido así desde hace mucho tiempo, desde que yo me convertí en químico clínico y parece ser un problema crónico que estalla con cada nueva generación de científicos del laboratorio. Estudiamos el uso y la interpretación de las estadísticas en los estudios de comparación de método hace más de treinta años y las lecciones que aprendimos aún sirven hoy en día. Nuestra intención aquí es brindar breves lineamientos —las estadísticas como tal, se revisan con más detalle en el siguiente capítulo.

Graficando los datos

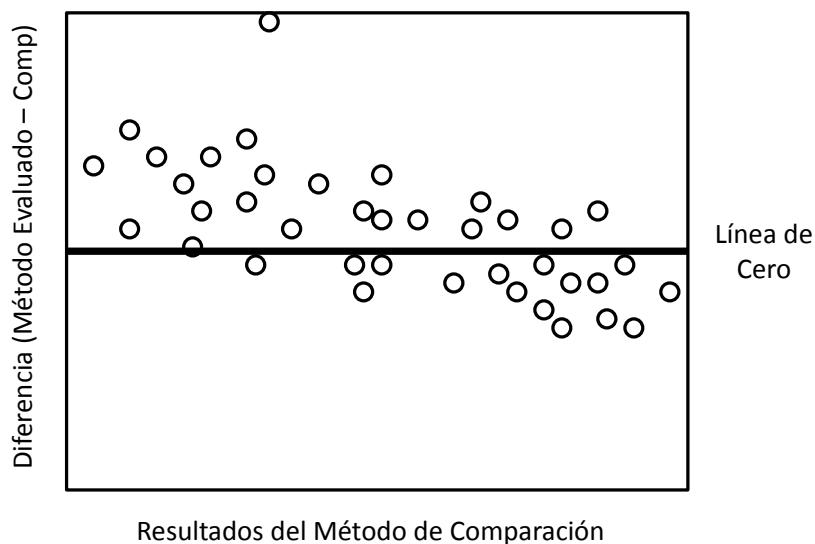
La técnica fundamental de análisis de datos consiste en graficar los resultados de la comparación e inspeccionar visualmente los datos. Idealmente, esto debería ser realizado *en el momento en que los datos son recolectados* para identificar resultados discrepantes que complicarán el análisis de datos. Cualquier espécimen de paciente con resultados discrepantes entre el método analítico a evaluar y el de comparación debería ser analizados nuevamente para confirmar que las diferencias sean reales y no equivocaciones al registrar los valores o mezclar los especímenes.

Si se espera que los dos métodos concuerden de manera absoluta, el gráfico inicial podría ser un “Gráfico de Diferencias” que muestre la diferencia entre los resultados del método evaluado y los resultados de método de comparación en el eje de las ordenadas “Y” contra el resultado obtenido a partir del método de comparación en el eje de las abscisas

“X”, tal como se muestra en la figura. Estas diferencias deberían ubicarse, alrededor de la línea de “cero diferencias”, la mitad arriba y la mitad abajo. Las diferencias grandes van a sobresalir y nuestra atención estará dirigida a aquellos especímenes cuyos resultados necesitan ser confirmados por mediciones repetidas. Busque cualquier punto atípico que no caiga en el patrón general de los otros puntos obtenidos a partir de los datos. Por ejemplo, hay un punto sospechoso en el “Gráfico de Diferencias”. Note también que los puntos tienden a dispersarse sobre la línea a concentraciones bajas y bajo la línea a concentraciones altas, sugiriendo que podría haber errores sistemáticos constantes y/o proporcionales.

Para situaciones donde no esperamos una concordancia absoluta entre los métodos, una concordancia absoluta, por ejemplo en análisis enzimáticos que tienen diferentes condiciones de reacción, deberíamos aplicar un “Gráfico de Comparación” que muestre el resultado del método a ser evaluado en el eje de las ordenadas “Y” contra el resultado obtenido a partir del método de comparación en el eje de las abscisas “X”, como se muestra en la segunda figura abajo.

“Gráfico de Diferencia”



Conforme se acumulan los puntos, debería dibujarse una línea visual que representa el mejor ajuste para mostrar la relación general entre los métodos y ayudar a identificar resultados discrepantes. Nuevamente, el propósito de ésta inspección gráfica inicial de datos es identificar resultados discrepantes para reanalisar especímenes mientras están frescos y aún disponibles.

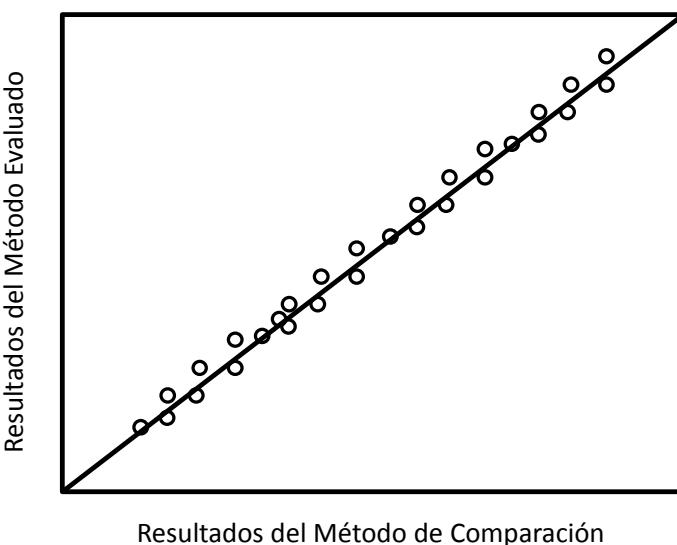
Sin embargo, éste tipo de gráfico generalmente es ventajoso para mostrar el rango analítico de datos, la linealidad de la respuesta a lo largo del rango, y la relación general entre métodos como se muestra por el ángulo de la línea y su intersección con el eje de las ordenadas.

Calcule las estadísticas apropiadas

Mientras que los “Gráficos de Diferencias” y los “Gráficos de Comparación” proveen impresiones visuales de los errores analíticos entre el método analítico evaluado y el método de comparación, las estimaciones numéricas de estos errores pueden ser obtenidas a partir de cálculos estadísticos. Recuerde que el significado secreto, más profundo y escondido de la validación de métodos es el análisis de errores. Necesita saber qué tipos de errores están presentes y qué tan grandes podrían ser. Los cálculos estadísticos pondrán números más exactos a sus impresiones visuales de los errores.

Dado que el propósito del experimento de comparación de métodos es valorar la inexactitud, las estadísticas calculadas deberían proveer información acerca del error sistemático en concentraciones importantes que representan niveles de decisión médica. Además, sería útil conocer la naturaleza constante o proporcional de ese error (revise, *El plan experimental* para definiciones de errores constantes y proporcionales). Esta última información es útil para determinar la causa o fuente del error sistemático y valorar la posibilidad de mejorar el desempeño del método. Para una discusión adicional sobre el uso e interpretación de las estadísticas, vea el capítulo 12 y los archivos PDF de la referencia [3].

“Gráfico de Comparación”



Resultados del Método de Comparación

Para resultados de comparación que cubran un amplio rango analítico, por ejemplo, glucosa o colesterol, se prefieren estadísticas de regresión lineal. Estas estadísticas permiten la estimación del error sistemático a más de una concentración que represente o este próxima a niveles de decisión médica para evaluar la aceptabilidad del método, y también brindan información sobre la naturaleza proporcional o constante del error sistemático para poder así valorar las fuentes posibles de errores. Los programas estadísticos proveen comúnmente cálculos de regresión lineal o análisis de cuadrados mínimos para la pendiente (b) y la intersección-Y (a) de la línea de mejor ajuste y la desviación estándar de los puntos alrededor de esa línea ($S_{x/y}$). El error sistemático (ES) a una concentración que representa un nivel de decisión médica dado (X_c) se determina entonces calculando el valor “Y” correspondiente (Y_c) a partir de la regresión lineal, y luego se toma la diferencia entre Y_c y X_c , como sigue:

$$Y_c = a + bX_c$$

$$ES = Y_c - X_c$$

Por ejemplo, en el estudio de comparación de colesterol, donde la línea de regresión es $Y = 2.0 + 1.03X$, esto es, el intercepto-Y es 2.0 mg/dL y la pendiente (b) es 1.03, el valor Y_c que corresponde a un nivel crítico de decisión de 200 (X_c) sería 208 ($Y_c = 2.0 + 1.03*200$), lo que significa que hay un error sistemático de 8 mg/dL (208 – 200) para un nivel de decisión médica de 200 mg/dL.

También es común calcular el coeficiente de correlación, “r”, que es útil sobre todo para evaluar si el rango de datos es lo suficientemente amplio como para brindar buenas estimaciones de la pendiente y la intersección-Y más que para evaluar la aceptabilidad del método [3]. Cuando r es 0.99 o mayor, los cálculos de regresión lineal simple deberían proveer estimaciones confiables de la pendiente y la intersección-Y. Si “r” es menor a 0.99, sería mejor recolectar datos adicionales para expandir el rango de concentraciones evaluadas. Considere usar cálculos de estadísticos “t” para estimar el error sistemático en la media de los datos, o utilizar cálculos de regresión más complicados que son apropiados para un rango de datos más estrecho.

Para resultados de comparación que cubran un rango analítico estrecho, por ejemplo, sodio o calcio, etc., usualmente es mejor calcular la diferencia promedio entre los resultados, que es lo mismo que la diferencia entre los promedios de los dos métodos, que también se llama comúnmente sesgo. Este sesgo calculado está disponible típicamente en programas estadísticos que proveen cálculos de estadísticos “t” para datos apareados. Los cálculos también incluyen la desviación estándar de las diferencias (S_{dif}), la cual describe la distribución de estas diferencias entre métodos, y un valor “t”, que puede ser usado para interpretar

si los datos son suficientes para concluir que realmente hay un sesgo o diferencia entre los métodos. Los capítulos 8 y 12 discuten en detalle estas estadísticas.

Por ejemplo, si el promedio de 40 análisis de sodio por el método analítico evaluado es de 141.0 mmol/L y el promedio para los mismos especímenes por el método de comparación es de 138.5, entonces el error sistemático promedio, o sesgo, es de 2.5 mmol/L (141.0 – 138.5). El signo algebraico de éste sesgo es útil para mostrar qué método arroja resultados más bajos o menores, pero es el valor absoluto de la diferencia el que es importante para evaluar la aceptabilidad del método.

Criterios para un desempeño aceptable

El juicio de aceptabilidad depende de qué cantidad de error analítico es permitida sin afectar o limitar el uso y la interpretación de resultados analíticos individuales. Esto es complicado por el hecho de que cualquier resultado analítico individual también está sujeto a error aleatorio, así que el conjunto o error total (ET) está compuesto del error sistemático (ES) más el error aleatorio (EA). Este “Error Total” puede ser calculado como sigue:

$$ET_{\text{calc}} = ES + EA$$

$$ET_{\text{calc}} = \text{Sesgo}_{\text{media}} + 3S_{\text{media}}$$

donde S_{media} es el estimado de la desviación estándar del método en evaluación obtenido a partir del experimento de replicación y $\text{Sesgo}_{\text{media}}$ es la diferencia promedio o la diferencia entre los promedios de los cálculos del estadístico “t” o la diferencia entre $Y_c - X_c$ donde $Y_c = a + bX_c$ es obtenida a partir de las estadísticas de regresión. El desempeño del método es aceptable cuando el “Error Total” calculado (ET_{calc}) es menor que el “Error Total Permitido” (TE_a , por sus siglas en Inglés). Recuerde que los criterios de CLIA de competencia analítica para un desempeño aceptable están expresados en forma de Errores Totales Permitidos y son un buen punto de inicio para asentar requisitos para la calidad analítica.

Se pueden hacer evaluaciones similares de aceptabilidad usando el Cuadro de Decisión de Métodos [6], que se describe en el capítulo 16. Esta gráfico le permite graficar el $\text{Sesgo}_{\text{media}}$ en el eje de las ordenadas “Y” y la S_{media} en el de las abscisas “X”, y luego evaluar la aceptabilidad del método evaluado a partir de la ubicación del “punto operativo” relativo a las líneas para diferentes criterios de error total en el grafico.

Verificación del Reporte del Fabricante

CLIA no requiere estrictamente que el laboratorio evalúe la *aceptabilidad* de ningún sesgo o error sistemático observado relativo a cualquier estándar de calidad, incluso a su propio criterio para un desempeño aceptable en los exámenes de esquemas de valuación de la competencia. En su lugar, CLIA enfatiza la *verificación de las especificaciones declaradas por los fabricantes para exactitud* (o inexactitud, sesgo, error sistemático). Si un fabricante establece en sus especificaciones que no hay sesgo, el estadístico “t” revelará si los datos del laboratorio confirman o no las especificaciones declaradas por el fabricante. Si el valor calculado del estadístico “t” para datos apareados es menor que el valor crítico de “t”, entonces no se ha observado un sesgo real o estadísticamente significativo. Si un fabricante reporta un sesgo determinado, entonces es útil determinar el intervalo de confianza para su estimación del sesgo y compararlo con el sesgo declarado por el fabricante en sus especificaciones de desempeño. Si el sesgo declarado por el fabricante en sus especificaciones de desempeño cae dentro de los límites de confianza alto y bajo del sesgo observado, entonces se confirma la especificación declarada por el fabricante; si esto no es así, la especificación del fabricante no ha sido verificada. Note una simplificación importante: si el sesgo observado por usted es menor que el sesgo declarado por el fabricante en sus especificaciones de desempeño, habrá verificado las especificaciones de desempeño del fabricante y no son necesarios más cálculos estadísticos.

Por ejemplo, suponga que un conjunto de 41 comparaciones de pacientes para colesterol dio una media de 200 mg/dL, un sesgo de 2.5 mg/dL, una S_{dif} de 4.3 mg/dL y un valor de “t” de 2.58. Si el fabricante declara en sus especificaciones que no hay sesgo, entonces aplique el estadístico “t” para asegurarse de que esto es verdad. Busque el valor crítico de “t” para 40 grados de libertad ($N - 1$) y una $p=0.05$ en la página 95. Compare el valor de “t” calculado de 2.58 con el valor crítico de “t” de 2.02 ($L=40$, $p=0.05$). Como el valor de “t” calculado (2.58) es mayor que el valor crítico de “t” (2.02), se ha observado un sesgo estadísticamente significativo o real. Sus resultados experimentales son diferentes y NO verifican las especificaciones de desempeño declaradas por el fabricante para exactitud que establecen un de cero sesgo.

Note, sin embargo, que el método aún podría desempeñarse aceptablemente dado que el sesgo real de 2.5 mg/dL es pequeño comparado con el error total permitido de CLIA del 10%.

Si el fabricante reporta un sesgo de 1.5 mg/dL, entonces sería útil calcular el intervalo de confianza alrededor del sesgo que usted ha estimado de 2.5 mg/dL. El intervalo se calcula como sigue: $2.5 \text{ mg/dL} + t(S_{\text{dif}})/N^{1/2}$, donde el término que acompaña al + es $2.02*4.3/6.4$ o 1.4 mg/dL. Su estimación del sesgo está entre 1.1 y 3.9 mg/dL, lo que incluye el sesgo declarado por el fabricante en sus

especificaciones que es de 1.5 mg/dL, por lo que sus resultados experimentales han confirmado las especificaciones de desempeño declaradas por el fabricante para exactitud.

Estudios mínimos recomendados

Seleccione 40 especímenes de pacientes para cubrir el rango reportable entero del método. Analice 8 especímenes por día dentro de 2 horas por el método analítico evaluado y el método de comparación. Grafique los resultados inmediatamente en un “Gráfico de Diferencias” y busque discrepancias; reanalice cualquier espécimen que dé resultados discrepantes para eliminar valores atípicos e identificar interferencias potenciales. Continúe el experimento por 5 días si no se observan resultados discrepantes. Continúe otros 5 días si se observan discrepancias durante los primeros 5 días. Prepare un “Gráfico de Comparación” a partir de todos los datos para evaluar el rango, valores atípicos y la linealidad. Calcule el coeficiente de correlación y si “r” es 0.99 o mayor, calcule las estadísticas aplicando un modelo de regresión lineal simple y estime el error sistemático a concentraciones que representen o estén próximas a niveles de decisión médica. Si $r < 0.975$, estime el sesgo en la media de los datos a partir del estadístico “t” o, alternativamente, a partir de técnicas más sofisticadas de regresión [4]. Use el Cuadro de Decisión de Método para combinar los estimados del error sistemático y el error aleatorio y haga una evaluación del Error Total observado para el método.

Direcciones Futuras

El número recomendado de muestras de pacientes para el experimento de comparación de métodos parece reducirse con los años. Los lineamientos actuales de CLIA dan 20 como “regla general” para los experimentos de validación de métodos [7]. La guía EP15 del CLSI para la verificación por parte del usuario de la “veracidad” o sesgo también recomienda un mínimo de 20 muestras de pacientes. Además, la EP15 incluye una metodología alternativa que hace uso de un mínimo de 2 materiales de referencia con valores asignados que deben ser medidos por duplicado de 3 a 5 corridas (ver capítulo 19). Los materiales de referencia aquí podrían incluir Materiales de Referencia Certificados, tales como los obtenidos del E.U.A. *National Institute of Standards and Technology* (NIST, por sus siglas en Inglés), pero también podrían ser muestras de exámenes de esquemas de evaluación de la competencia, materiales de control de la calidad o materiales de comparación de grupos pares con valores asignados. La adopción de esta metodología alternativa que usa materiales de referencia ciertamente disminuye los obstáculos y reduce los esfuerzos necesarios para verificar las especificaciones de desempeño establecidas por los fabricantes. No debe usted olvidar que el verdadero propósito de la validación de métodos es asegurarse

de que su laboratorio está generando resultados analíticos correctos para los pacientes que concurren al laboratorio, no solamente mantener la conformidad con los requisitos regulatorios. Es importante ser minucioso y asegurarse que su método se desempeña de forma aceptable y que provee la calidad analítica requerida para el cuidado del paciente.

Referencias

1. CLSI EP9-A2. Method Comparison and Bias Estimation using Patient Samples. Clinical and Laboratory Standards Institute, Wayne, PA, 2002.
2. Hyltoft-Petersen P, Stockl D, Blaaberg O, Pedersen B, Birkemose E, Thienpont L, Flensted Lassen J, Kjeldsen J. Graphical interpretation of analytical data from a comparison of a field method with a reference method by use of difference plots. *Clin Chem* 1997;43:2039-2046.
3. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method-comparison studies. *Clin Chem* 1973;19:49-57.
4. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison studies. *Clin Chem* 1979;25:432-438.
5. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974;20:825-833.
6. Westgard JO. A method evaluation decision chart (MEDx chart) for judging method performance. *Clin Lab Science* 1995;8:277-283.
7. CLIA Brochure: Validation of Performance Specifications.
http://www.cms.hhs.gov/CLIA/05_CLIA_Brochures.asp
8. CLSI EP15-A2. User Verification of Performance for Precision and Trueness. Clinical and Laboratory Standards Institute, Wayne, PA, 2006.

Referencias Online

Herramientas para el análisis de datos de los estudios de validación de métodos.
<http://www.westgard.com/mvtools.html>

Uso e interpretación de las estadísticas que generalmente se aplican en los estudios de validación de métodos.

<http://www.westgard.com/downloads/>

Problema — Datos de Validación de un Método de Colesterol: Comparación de Métodos

Se ha llevado a cabo un estudio de comparación utilizando el método de Abell Kendal como método de comparación. Grafique los datos de comparación, obtenga las estadísticas de regresión y determine el valor del coeficiente de correlación, "r". (Respuestas en la página 306).

Muestra	Abell-Kendal Método de Comparación (X)	Método Evaluado (Y)
1	217	203
2	224	213
3	298	279
4	172	160
5	198	189
6	274	262
7	253	238
8	197	275
9	266	211
10	151	149
11	166	151
12	163	151
13	215	205
14	151	133
15	263	252
16	226	212
17	239	226
18	162	147
19	253	235
20	159	157
21	261	250
22	247	231
23	261	238
24	184	179
25	295	284
26	250	232
27	201	196
28	209	212
29	286	275
30	158	142
31	288	281
32	161	145
33	183	171
34	252	239
35	285	277
36	194	190
37	240	230
38	180	177
39	297	275
40	210	188



Munster's Monsters!
Sebastian Munster, 1598

Todos nosotros debemos lidiar con monstruos, tales como esas feroces criaturas de mar que decoraban muchos de los mapas del siglo XVI. Uno de nuestros monstruos modernos que debemos enfrentar a diario en los laboratorios es llamado “estadística”. Mucho atemorizan las estadísticas, pero como los Munster’s Monsters, las estadísticas son criaturas imaginarias que podemos superar con educación y entrenamiento. Aprender a usar las estadísticas de los estudios de validación de métodos requiere que usted supere sus temores y adopte a las estadísticas como una herramienta práctica que le ayudará a resumir y comprender el significado de los números que surgen de estos estudios. El uso apropiado de las estadísticas es particularmente importante para comprender los resultados de los estudios de comparación de métodos.

12: ¿Cómo usar la estadística para estimar errores analíticos?

Este capítulo provee más detalles acerca del análisis e interpretación de los datos de un experimento de comparación de métodos. Hace uso de datos simulados para demostrar el comportamiento de los parámetros estadísticos en respuesta a diferentes tipos de errores que podrían ser observados a partir de los datos.

Objetivos:

- Relacionar las estadísticas utilizadas para analizar los datos de un experimento de comparación de métodos con los tipos de errores analíticos que ocurren entre los métodos.
- Identificar las limitaciones del estadístico “t” y de las estadísticas de regresión.
- Formular una estrategia para realizar un análisis de datos apropiado de un experimento de comparación de métodos.

Materiales del capítulo:

- VM — Sentido Estadístico, Sensibilidad y Significatividad,
por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar los materiales.
- Preparar un conjunto de datos para demostrar el efecto del error proporcional sobre los estadísticos “t” y el gráfico de diferencias.
- Preparar un conjunto de datos para mostrar el efecto de trabajar con un rango estrecho de datos sobre los estadísticos de regresión y en el gráfico de comparación.
- Examinar un estudio de validación publicado en la literatura científica y criticar el uso y aplicación de las estadísticas.

Preguntas de Autoevaluación:

El siguiente resumen estadístico fue obtenido a partir de un experimento de comparación de métodos para glucosa:

$a = 5,23 \text{ mg/dL}$, $b = 0,999$, $Sy/x = 7,23 \text{ mg/dL}$, sesgo = $5,13 \text{ mg/dL}$, $Sdif = 7,23 \text{ mg/dL}$, $t = 8,03$, $r = 0,996$, $N = 128$.

- ¿Cuál es el error sistemático proporcional entre métodos?
- ¿Cuál es el error sistemático constante entre métodos?
- ¿Cuál es el error aleatorio entre métodos?
- ¿Por qué hay tan buena concordancia entre las estimaciones del error obtenido a partir de estadísticas de regresión y estadísticos “t”?
- ¿Es el error sistemático entre métodos estadísticamente significativo o “real”? ¿Qué le dice el coeficiente de correlación?

El siguiente resumen estadístico fue obtenido a partir de un experimento de comparación de métodos para nitrógeno ureico:

$N = 316$, $a = -0,31 \text{ mg/dL}$, $Sa = 0,23 \text{ mg/dL}$, $b = 1,032$, $Sb = 0,009$, $Sy/x = 0,97 \text{ mg/dL}$, $Sx = 13,2 \text{ mg/dL}$, $r = 0,997$, sesgo = $0,40 \text{ mg/dL}$, $Sdif = 1,08 \text{ mg/dL}$, $t = 6,58$.

- ¿Cuál es el error sistemático proporcional entre métodos?
- ¿Cuál es el error sistemático constante entre métodos?
- ¿Por qué es mejor usar las estadísticas de regresión para estimar los errores en lugar de usar un estadístico “t”?
- ¿Cuál es el intervalo de confianza del 95% para la intersección en “y”?
- ¿La intersección en “y” difiere significativamente del valor ideal de 0.0?
- ¿Cuál es el intervalo de confianza del 95% para la pendiente?
- ¿Difiere la pendiente significativamente del valor ideal de 1,00?
- ¿Qué le dice el coeficiente de correlación?

Validación de Métodos: Sentido Estadístico, Sensibilidad y Significatividad

James O. Westgard, PhD

Recuerde que el primer secreto de la validación de métodos — se trata totalmente de evaluación de errores. El segundo secreto es que las estadísticas son sólo herramientas para estimar el tamaño de esos errores. La **caja de herramientas de análisis de datos** le facilita calcular las estadísticas apropiadas y preparar los gráficos asociados utilizando calculadoras y herramientas para graficar *online*. Aún así, necesita ser cuidadoso al interpretar las estadísticas, particularmente para datos del experimento de comparación de métodos.

Hace muchos años, estudiábamos el uso y la interpretación de las estadísticas para el estudio de los datos provenientes de la comparación de métodos. Empleamos un modelo de simulación de datos para crear conjuntos de datos que tienen diferente tipo y magnitud de errores analíticos, luego calculamos las estadísticas de regresión, estadísticos “t” y los coeficientes de correlación para cada uno de los conjuntos de datos. Vimos qué estadísticas cambiaron conforme cambió el tipo y la magnitud de los errores analíticos en los conjuntos de datos.

Esos resultados fueron publicados en *Clinical Chemistry* [1] y el artículo original está disponible en Westgard Web en <http://www.westgard.com/downloads>. En 2008, *Clinical Chemistry* reconoció éste artículo como un “clásico para citar”. Si bien éste conocimiento ha estado disponible por alrededor de 30 años, no está siendo difundido en programas actuales de educación y entrenamiento. Muchos profesionales del laboratorio aún tienen gran dificultad para interpretar los estudios de comparación de métodos. La clave para hacer útiles a las estadísticas descansa en entender su sensibilidad a los distintos tipos de error que pueden estar presentes en los datos. **El sentido estadístico se relaciona con la sensibilidad de las estadísticas a los errores.**

Simulación de errores en resultados de examen

Los métodos antiguos de glucosa no tenían la especificidad de los métodos enzimáticos actuales. Debido a que varios estudios de comparación entre métodos específicos y no específicos de glucosa estaban apareciendo en la literatura en ese tiempo, la glucosa era un buen ejemplo para demostrar cómo los diferentes tipos de errores analíticos aparecen en los resultados estadísticos de los estudios de comparación de métodos.

Empezamos por construir un conjunto de datos de 41 especímenes que serían típicos de una población hospitalaria, como se muestra en la Tabla 1 colocando los resultados del método de referencia en la 1^a columna. En un mundo perfecto, el método de examen daría exactamente los mismos resultados, como se muestra en la 2^a columna. Para demostrar los efectos de los diferentes tipos de errores, se pueden crear conjuntos de datos adicionales de resultados de comparación manipulando los datos de referencia originales (1^a columna) de maneras específicas.

- El error aleatorio puede ser simulado añadiendo o sustrayendo alternativamente 5 mg/dL a cada punto de los datos de referencia (1^a columna), como se muestra en la tercera columna de resultados. Esto no es verdaderamente aleatorio, al menos no el error aleatorio Gaussiano normal esperado para un método analítico, pero será suficiente para demostrar el efecto de este tipo de error. Se pueden construir conjuntos de datos adicionales añadiendo ± 2 mg/dL y ± 10 mg/dL para demostrar el efecto generado por el cambio en la magnitud del error aleatorio.
- El error sistemático constante puede ser simulado añadiendo 10 mg/dL a cada punto del conjunto de datos de referencia (1^a columna), como se muestra en la cuarta columna. Se pueden construir datos adicionales añadiendo 2 mg/dL o 5 mg/dL para demostrar el efecto generado por el cambio en la magnitud del error sistemático constante.
- El error sistemático proporcional puede ser simulado multiplicando cada resultado del conjunto de datos de referencia (1^a columna) por 1,05. Se pueden construir datos adicionales usando factores de 1,02 y 1,10 para demostrar el efecto generado por el cambio en la magnitud del error sistemático proporcional.
- Se pueden simular combinaciones de errores aplicando dos o más de las operaciones de arriba, como se muestra en las columnas de datos 6, 7 y 8.

Tabla 1. Ejemplo de Simulación de datos glucosa

Referencia	Perfecto	Aleatorio	Constante	Proporcional	EA y ESC EA 5 y ESC 5	EA y ESP EA 5 y ESP 5	EA y ESC y ESP EA 5 y ESC 5 y ESP 5
40	40	35	45	38.00	45	33.25	42.75
60	60	65	65	57.00	65	61.75	61.75
80	80	75	85	76.00	85	71.25	80.75
90	90	95	95	85.50	95	90.25	90.25
100	100	95	105	95.00	105	90.25	99.75
110	110	115	115	104.50	115	109.25	109.25
120	120	115	125	114.00	125	109.25	118.75
125	125	130	130	118.75	130	123.50	123.50
130	130	125	135	123.50	135	118.75	128.25
135	135	140	140	128.25	140	133.00	133.00
140	140	135	145	133.00	145	128.25	137.75
145	145	150	150	137.75	150	142.50	142.50
150	150	145	155	142.50	155	137.75	147.25
155	155	160	160	147.25	160	152.00	152.00
165	165	160	170	156.75	170	152.00	161.50
170	170	175	175	161.50	175	166.25	166.25
175	175	170	180	166.25	180	161.50	171.00
180	180	185	185	171.00	185	175.75	175.75
185	185	180	190	175.75	190	171.00	180.50
190	190	195	195	180.50	195	185.25	185.25
195	195	190	200	185.25	200	180.50	190.00
200	200	205	205	190.00	205	194.75	194.75
205	205	200	210	194.75	210	190.00	199.50
210	210	215	215	199.50	215	204.25	204.25
215	215	210	220	204.25	220	199.50	209.00
220	220	225	225	209.00	225	213.75	213.75
225	225	220	230	213.75	230	209.00	218.50
230	230	235	235	218.50	235	223.25	223.25
235	235	230	240	223.25	240	218.50	228.00
240	240	245	245	228.00	245	232.75	232.75
245	245	240	250	232.75	250	228.00	237.50
250	250	255	255	237.50	255	242.25	242.25
260	260	255	265	247.00	265	242.25	251.75
270	270	275	275	256.50	275	261.25	261.25
280	280	275	285	266.00	285	261.25	270.75
290	290	295	295	275.50	295	280.25	280.25
300	300	295	305	285.00	305	280.25	289.75
320	320	325	325	304.00	325	308.75	308.75
340	340	335	345	323.00	345	318.25	327.75
380	380	385	385	361.00	385	365.75	365.75

Análisis estadístico de los datos simulados

Estos conjuntos de datos están sometidos a cálculos estadísticos de regresión (la pendiente, **b**; la intersección-Y, **a**; y la desviación estándar de los puntos alrededor de la línea de regresión, s_{yx}), el análisis del estadístico “t” para datos apareados (la diferencia promedio entre los métodos, sesgo; la desviación estándar de las diferencias, S_{dif} ; y el valor t calculado que se usa para determinar si el sesgo es estadísticamente significativo o “real”), y finalmente el coeficiente de correlación, **r**.

Tabla 2. Efecto de los errores analíticos en resultados calculados

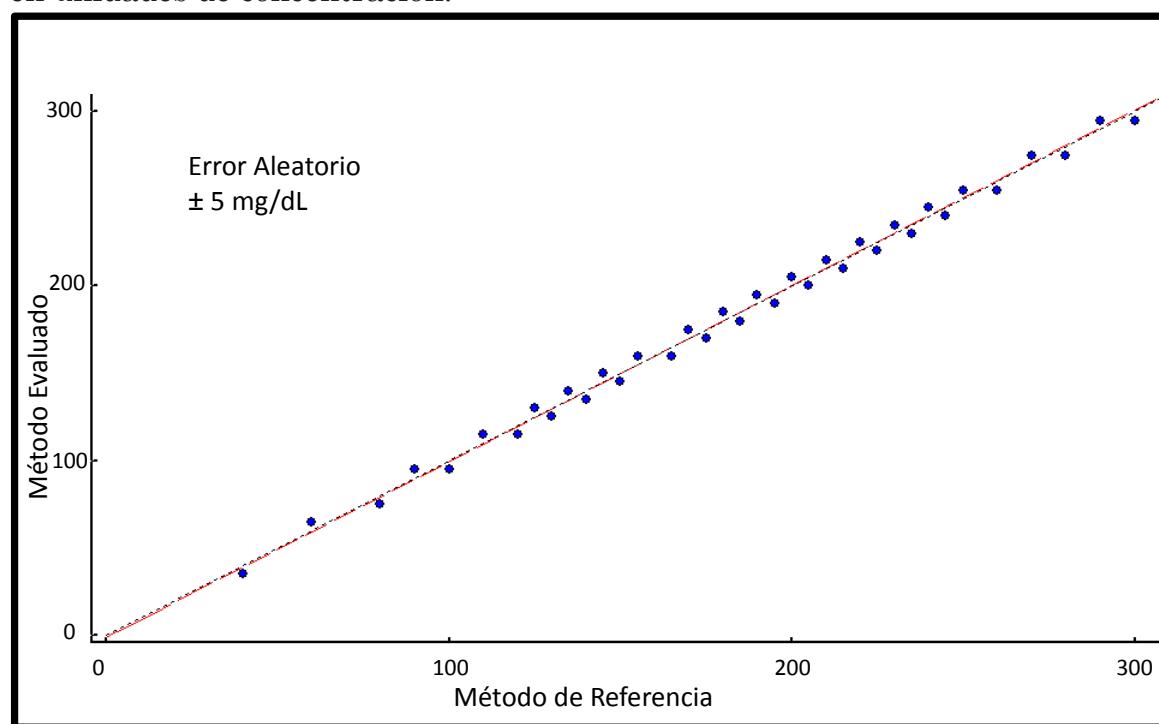
Conjunto de Datos	EA	ESC	ESP	Pendiente	Y-Intercepto	Sy/x	Sesgo	S_{dif}	t	r
1. Perfecto	0	0	0	1.000	0.00	0.00	0.00	0.00	Sin definir	1.000
2. EA 2	2	0	0	1.001	-0.02	2.00	0.05	2.00	0.16	0.999
3. EA 5	5	0	0	1.001	-0.04	5.00	0.12	5.00	0.16	0.993
4. EA 10	10	0	0	1.003	-0.08	10.00	0.24	10.00	0.16	0.986
5. ESC 2	0	2	0	1.000	2.00	0.00	2.00	0.00	Sin definir	1.000
6. ESC 5	0	5	0	1.000	5.00	0.00	5.00	0.00	Sin definir	1.000
7. ESC 10	0	10	0	1.000	10.00	0.00	10.00	0.00	Sin definir	1.000
8. ESP 2	0	0	0	0.980	0.00	0.08	2.29	1.18	12.43	1.000
9. ESP 5	0	0	5	0.950	0.00	0.08	5.72	2.95	12.42	1.000
10. ESP 10	0	0	10	0.900	-0.04	0.14	11.45	5.88	12.47	1.000
11. EA 5 + ESC 2	5	2	0	1.001	1.96	5.00	2.12	5.00	2.72	0.996
12. EA5 + ESC 5	5	5	0	1.002	4.92	5.02	5.10	5.03	6.50	0.996
13. EA5 + ESC 10	5	10	0	1.002	9.95	5.02	10.10	5.03	12.86	0.996
14. EA 10 + ESC 2	10	2	0	1.003	1.92	10.00	2.24	10.00	1.44	0.986
15. EA 10 + ESC 5	10	5	0	1.003	4.92	10.00	5.24	10.00	3.36	0.986
16. EA 10 + ESC 10	10	10	0	1.003	9.92	10.00	10.24	10.00	6.56	0.986
17. EA 2 + ESP 5	2	0	5	0.951	-0.02	2.00	5.67	3.53	10.27	0.999
18. EA 5 + ESP 5	5	0	5	0.951	-0.04	5.00	5.59	5.76	6.27	0.996
19. EA 10 + ESP 5	10	0	5	0.953	-0.08	10.00	5.47	10.38	3.38	0.985
20. EA 5 + ESP 10	5	0	10	0.900	0.23	5.01	11.28	7.85	9.09	0.996
21. EA 5 + ESP 25	5	0	25	0.751	-0.02	5.00	28.45	15.58	11.62	0.994
22. EA 5 +ESP 50	5	0	50	0.501	-0.04	5.00	57.02	30.17	12.10	0.986
23. ESC 2 + ESP 5	0	2	5	0.950	2.00	0.08	3.72	2.95	8.07	1.000
24. ESC 5 + ESP 5	0	5	5	0.950	5.00	0.08	0.72	2.95	1.55	1.000
25. ESC 10 + ESP 5	0	10	5	0.950	10.00	0.80	4.28	2.95	9.31	1.000
26. EA 2 + ESC 5 + ESP 10	2	5	10	0.905	4.99	2.03	6.37	6.28	6.50	0.999
27. EA 10 + ESC 2 + ESP 5	10	2	5	0.953	1.89	10.01	3.48	10.50	12.12	0.984
28. EA 5 +ESC 10 + ESP 2	5	10	2	0.981	9.96	4.99	7.84	5.18	9.69	0.996

Las estadísticas calculadas se muestran en la 2^a tabla, donde la primera columna a la izquierda (1^a columna) identifica el tipo y magnitud de los errores analíticos en el conjunto de datos (EA, error aleatorio; ESC, error sistemático constante; ESP, error sistemático proporcional) y las siguientes tres columnas muestran la magnitud de los errores. Cada fila provee los resultados estadísticos de la condición de error específico o del conjunto de condiciones de error. La comparación de las condiciones de error y los valores estadísticos nos permiten correlacionar el tipo y magnitud de los errores a los estimados de las estadísticas.

Sin errores. Los datos de comparación perfectos tienen exactamente los mismos valores para los métodos de examen y de referencia. En un gráfico de comparación, todos los puntos caerían exactamente en una línea dibujada en un ángulo de 45 grados y con la intersección-Y, justo en la intersección de los dos ejes “X” e “Y”, o sea cero de “x” y cero de “y”. Como se muestra en la tabla, los resultados estadísticos para ésta situación ideal muestran un valor de 1,00 para la pendiente y el coeficiente de correlación, y valores de 0,00 para todas las demás estadísticas, excepto el estadístico “t” que es indefinido (debido a que es una proporción de dos términos que son 0,00). ***En un mundo perfecto, las estadísticas tienen valores ideales de 1,00 (pendiente, coeficiente de correlación) y 0,0 (intersección-Y, $S_{y/x}$, sesgo, S_{dif}). Desviaciones de estos valores ideales son indicadores de error.***

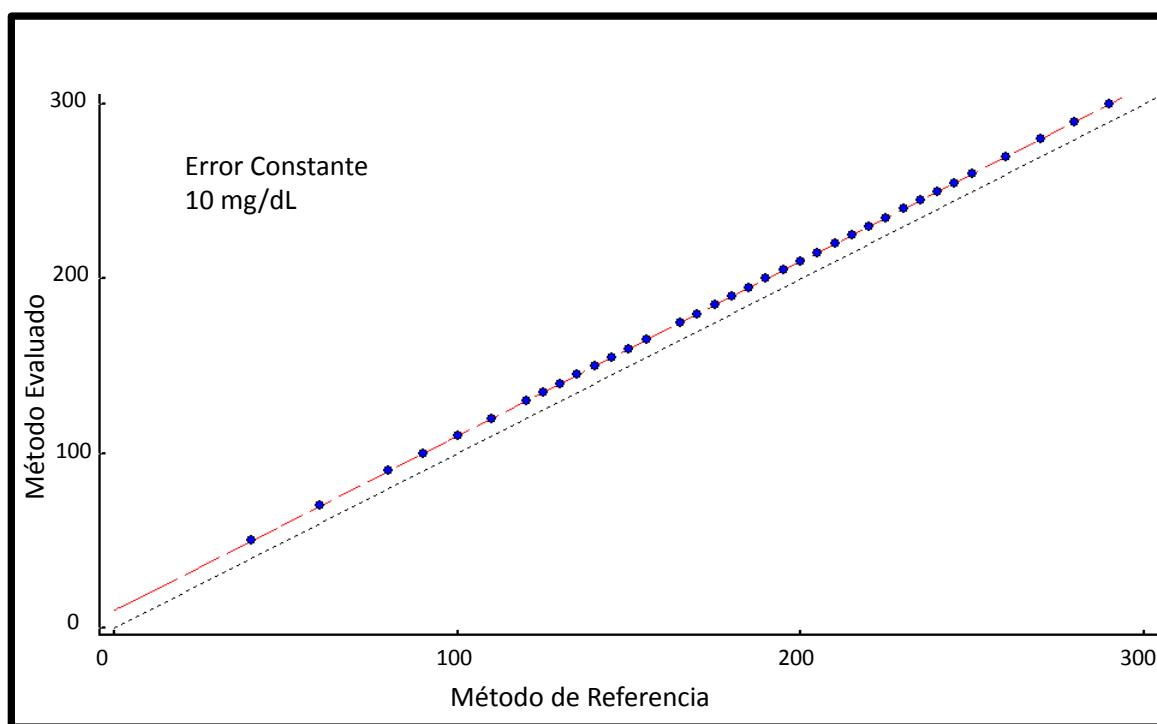
Error aleatorio.

El efecto del error aleatorio se muestra en ésta figura para los datos donde 5 mg/dL han sido añadidos y sustraídos alternativamente a los valores de referencia. El error aleatorio se manifiesta en el gráfico como dispersión en los puntos alrededor de la línea de regresión. Los cálculos estadísticos en la Tabla 2 muestran cambios mínimos en la pendiente, la intersección o el sesgo, pero las $S_{y/x}$ y S_{dif} reflejan directamente el tamaño del error aleatorio. El coeficiente de correlación disminuye conforme se incrementa el error aleatorio, pero los cambios en “r” son pequeños y no proveen un estimado directo del error aleatorio en unidades de concentración.



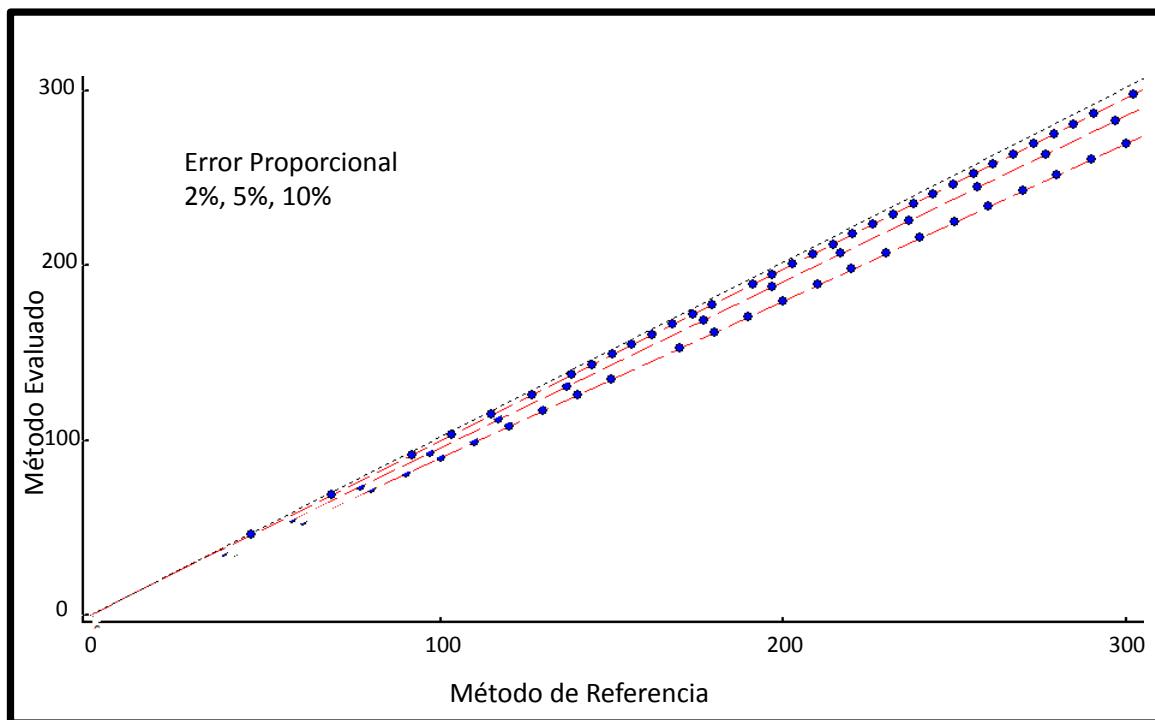
Error Sistemático Constante.

El efecto del error sistemático constante se muestra en la siguiente figura, donde la línea de regresión no pasa ya por el origen. En este caso, se añadieron 10 mg/dL para crear un error sistemático constante y la magnitud del error está correctamente estimada tanto por la intersección-Y de la regresión como por el sesgo del análisis del estadístico “t”. Note que no hay cambios en el coeficiente de correlación.



Error Sistemático Proporcional.

El efecto del error sistemático proporcional es mostrado por tres situaciones donde los resultados de los exámenes son menores que los de referencia por un 2%, 5% y un 10%. El error sistemático proporcional cambia la pendiente de la línea de regresión y la magnitud exacta del error se estima por la pendiente del análisis de regresión. Note que el error sistemático proporcional no afecta la intersección-Y, $S_{y/x}$ y “r”, pero causa cambios en el sesgo y la S_{dif} en análisis del estadístico “t”.



Sensibilidad de la estadística a los tipos de error

El comportamiento de los distintos estadísticos respecto a los diferentes tipos de error se resume en la tabla mostrada aquí. El error aleatorio se refleja en cambios en $S_{y/x}$, S_{dif} y r . Los errores sistemáticos constantes se muestran en la intersección en "y", y en el sesgo. El error sistemático proporcional puede ser estimado de mejor manera por la desviación de la pendiente de la ideal, pero también causa cambios en el sesgo y la DE_{dif} del análisis del estadístico t . Esto es un problema y la razón del signo de interrogación en la tabla de abajo. ¡El error sistemático proporcional confunde la interpretación de los estadísticos t ! También existe un problema con el coeficiente de correlación porque responde sólo al error aleatorio, no a errores sistemáticos, que son los errores de interés en el experimento de comparación de métodos. ¡Puede tener un coeficiente de correlación ideal incluso si un método es inexacto!

Sensibilidad de la Estadística	EA	ESC	ESP
Regresión			
Pendiente, b	No	No	Si
Y-Intercepto, a	No	Si	No
DE alrededor de la línea, $S_{y/x}$	Si	No	No
Prueba - T			
Sesgo	No	Si	Si?
S_{dif}	Si	No	Si?
Coeficiente de Correlación			
r	Si?	No?	No?

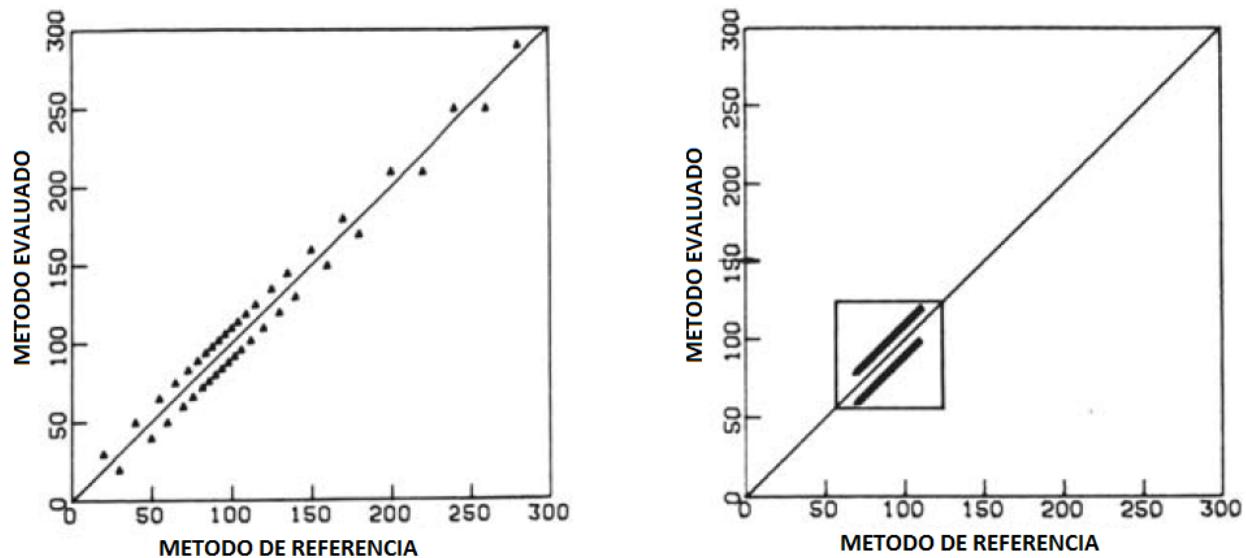
Estimación del error aleatorio.

De la tabla de sensibilidad, es aparente que hay tres estadísticas que responden al error aleatorio. Mientras que éste error fue introducido en el método de examen por simulación de datos, las estadísticas a menudo reflejan el error aleatorio de ambos métodos y a veces hasta incluyen dispersión adicional causada por diferencias en la especificidad entre los dos métodos que están siendo comparados. Por lo tanto, el estimado de la precisión del experimento de replicación es importante todavía para caracterizar el desempeño de un método individual.

Note que la S_{dif} también está influenciado por el error sistemático proporcional, lo que significa que ésta estadística no provee un estimado específico de

error aleatorio. Las otras dos estadísticas — $S_{y/x}$ y “r” — son sensibles sólo al error aleatorio, pero difieren en sus unidades y valores numéricos. Es más útil estimar el error aleatorio como desvío estándar, en unidades de concentración como nos ofrece el $S_{y/x}$, que en los números adimensionales provistos por el coeficiente de correlación. Un valor de 5 mg/dL para $s_{y/x}$ puede ser fácilmente interpretado en términos de diferencias esperadas, por ejemplo, 95% de las diferencias estarán dentro de 2 veces el valor de $S_{y/x}$. Un valor de 0,996 para “r” no provee un estimado del tamaño de los errores aleatorios en una forma significativa como un término de desvío estándar.

Una limitación mayor de “r” es que depende del rango analítico cubierto por los datos. Por ejemplo, en los dos gráficos mostrados aquí, el error aleatorio es el mismo, ± 10 mg/dL, aunque los valores para r son muy diferentes, 0,986 vs 0,764. El gráfico a la izquierda muestra el amplio rango de datos que se esperarían de una población de un hospital, mientras que el gráfico a la derecha muestra el rango estrecho esperado de una población saludable.



Rango	0 a 300 mg/dL
Error Aleatorio	10 mg/dL
Coeficiente de Correlación	0.986

Rango	70 a 110 mg/dL
Error Aleatorio	10 mg/dL
Coeficiente de Correlación	0.764

El coeficiente de correlación es de hecho sensible al error aleatorio entre métodos (la dispersión en la dirección “y”) *relativo al rango de datos analíticos en la dirección “x”*. Un alto coeficiente de correlación significa que la dispersión en “y” es pequeña comparada con la distribución en “x”. Aunque que éste comportamiento hace que coeficiente de correlación tenga poco valor para estimar

errores analíticos, sí de los valores obtenidos para la pendiente y el intercepto-Y durante la comparación de métodos, una medida sobre confiabilidad de los valores obtenidos para la pendiente y el intercepto-Y durante la comparación de métodos, esto es, un valor alto de “r” significa que los datos cubren un amplio rango relativo a la dispersión entre métodos, haciendo que la línea que pasa a través de los datos esté bien definida.

Estimación del error sistemático constante.

La intersección en “y” de la regresión y el sesgo del estadístico “t” son las mejores estadísticas para estimar el error sistemático constante. Ambos dan estimados en unidades de concentración y proveen valores similares cuando está ausente el error sistemático proporcional. Considerando que el error sistemático proporcional afecta al sesgo del análisis del estadístico “t”, este estadístico no ofrece un estimado específico del error sistemático constante. En su lugar provee un estimado general del error sistemático que es confiable sólo en la media de los datos. Recuerde que el sesgo es calculado como la diferencia entre la media de los resultados del método de examen menos la media de los resultados del método de comparación, que es lo mismo que el promedio de las diferencias de todos los especímenes individuales.

Estimación del error sistemático proporcional.

La pendiente de la regresión, así como el sesgo y la S_{dif} del estadístico “t” son todos sensibles al error sistemático proporcional. El hecho de que tanto el sesgo como la S_{dif} responden demuestra que éstos estadísticos no pueden proveer un estimado específico de error sistemático proporcional. Ambos pueden ser desorientadores debido a que también responden a otros tipos de error, esto es, el sesgo responde también al error sistemático constante y la S_{dif} responde también a errores aleatorios. Sería mejor evitar usar estadísticos “t” cuando está presente el error sistemático proporcional.

La regresión provee el mejor estimado del error sistemático proporcional. La diferencia entre la pendiente y su valor ideal de 1,00, expresado como porcentaje, describe el error sistemático proporcional en la forma más útil. Por ejemplo, una pendiente observada de 0,95 indica un error sistemático proporcional del 5,0%.

Dándole Sentido a las Estadísticas

Coeficiente de correlación.

El coeficiente de correlación provee información sólo sobre el error aleatorio, incluso si el objetivo de un estudio de comparación de métodos es estimar el error sistemático. Por lo tanto, **el coeficiente de correlación es de poco valor para estimar errores analíticos en un experimento de comparación de métodos**. Sin embargo, dado que “r” es sensible al rango de datos recolectados, “r” es útil para medir la **confiabilidad de las estadísticas de regresión**. [¿No es maravilloso? Una limitación puede ser convertida en una herramienta útil una vez que el comportamiento está bien entendido].

Estadísticos “t”.

Las estimaciones de los errores podrían ser confundidas por la presencia de error sistemático proporcional. Hay dos casos donde las estimaciones de error sistemático serán confiables: (1) si el error sistemático proporcional está ausente, entonces el error sistemático es constante a lo largo de todo el rango de concentración; (2) si la media de los resultados de los pacientes es cercana al nivel de decisión médica de interés, entonces el estimado general del error sistemático constante y proporcional es confiable en la media de los datos, pero ese estimado de error sistemático no debería ser extrapolado a concentraciones de otros niveles de decisión médica.

- La inspección de un gráfico con valores del método de ensayo en el eje “y”, y valores del método de comparación en el eje de las “x” usualmente revelará la presencia de error sistemático proporcional.
- Si NO está presente el error sistemático proporcional, entonces el estimado del error constante podría aplicar a todo el rango de datos estudiado, pero aún así sería mejor restringir la interpretación de los niveles de decisión médica cerca de la media de los datos.
- Si está presente el error sistemático proporcional, sería mejor usar estadísticas de regresión para estimar el error sistemático a niveles de decisión de interés.

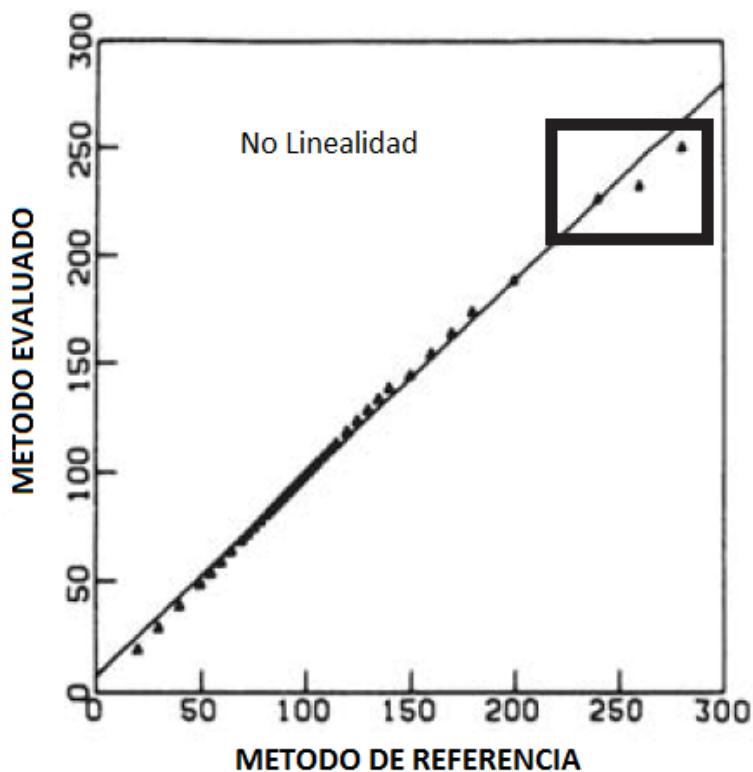
Estadísticas de regresión.

Es ideal tener tres parámetros estadísticos que puedan estimar cada uno un tipo diferente de error. El error sistemático proporcional puede ser estimado a partir de la pendiente, el error sistemático constante a partir de la intersección-Y, y el error aleatorio (entre métodos) a partir de la desviación estándar alrededor de la línea de regresión. El error sistemático puede ser estimado a

cualquier concentración usando la ecuación de regresión, esto es, $Y_c = a + bX_c$, donde X_c es la concentración que está próxima o representa un nivel de decisión médica y Y_c es el mejor estimado de la concentración por el método de examen. La diferencia entre Y_c y X_c es el error sistemático a esa concentración crítica, esto es, $Y_c - Y_x = ES$.

Las estimaciones de los errores a partir de las estadísticas de regresión no serán confiables a menos que los datos satisfagan ciertas condiciones y asunciones.

- Se asume la linealidad, por lo que debe inspeccionar un gráfico de resultados de comparación para asegurarse de que hay una relación lineal entre los dos métodos. Por ejemplo, en el gráfico de abajo se muestra el efecto de la no-linealidad en la línea de regresión. Aunque es obvio que hay algo de no-linealidad en el extremo superior, los cálculos para la regresión lineal determinarán la mejor línea recta a través de todos los datos. En éste caso, los puntos en el extremo superior llevarán la línea hacia abajo, haciendo que la pendiente baje y aumentando el valor del intercepto-Y. Las estimaciones de error sistemático proporcional y error sistemático constante serán corrompidas por cualquier falta de linealidad en los datos.



- ¡Los valores atípicos pueden causar un problema similar! Uno o dos puntos al final de la línea pueden ejercer una influencia indebida arrastrando la línea hacia ellos y afectando tanto la pendiente como la intersección-Y. El remedio es nuevamente inspeccionar el gráfico de los datos para asegurarse de que no hay valores atípicos.
- Un rango estrecho de datos también es un problema porque una línea no puede ser bien definida por una nube de puntos de datos. Mientras que ésta situación puede ser reconocida en ocasiones a partir de la observación de un gráfico de datos, la mejor alerta está dada por el coeficiente de correlación. Un valor bajo de "r", menor de 0.99 según algunas referencias o menor de 0.975 según otras [2], indica que las estimaciones de la pendiente y la intersección podrían ser afectadas por la dispersión de los datos. Un remedio es utilizar técnicas de regresión más sofisticadas, tales como regresión de Deming o regresión de Passing-Bablok. Un remedio más simple es utilizar una estimación del sesgo obtenida a partir del estadísticos "t" e interpretar los datos en la media de los resultados de los pacientes (asumiendo que la media es cercana al nivel de decisión médica de interés).

Significatividad estadística vs clínica.

No hemos incluido el valor "t" en la discusión hasta ahora porque ¡no provee un estimado de los errores! Éste estadístico es un "examen de significatividad" que es principalmente útil para decidir si se han juntado suficientes datos para demostrar que existe una diferencia. Si el valor "t" calculado es mayor que el valor "t" crítico (que es 2,02 para el ejemplo de conjuntos de datos con 41 puntos), se dice que el sesgo observado es estadísticamente significativo, que en términos prácticos significa "real". Si el valor "t" calculado es menor que el valor crítico de "t", entonces los datos no son suficientes para demostrar que existe un "sesgo estadísticamente significativo" entre los valores obtenidos a partir del método en evaluación y el método de comparación.

Desde mi perspectiva, ésta información sobre la significatividad estadística es de importancia secundaria. El evaluación sobre la aceptabilidad del método debe hacerse sobre la significatividad clínica, no sobre la significatividad estadística. Un error puede ser estadísticamente significativo, esto es, real, aunque tan pequeño que no sea clínicamente importante. Por otro lado, un error puede ser grande y clínicamente importante, aunque los datos no sean suficientes para demostrar que es estadísticamente significativo.

- Recuerde que el valor "t" se calcula como $t = \text{sesgo} * (N^{1/2}/S_{\text{dif}})$, lo que muestra que es una proporción del error sistemático (sesgo) dividido por el error aleatorio (S_{dif}) multiplicado por la raíz cuadrada del número

de muestras apareadas ($N^{1/2}$). Esto es análogo a la ecuación para el pH sanguíneo, donde el pH es una función de la proporción del bicarbonato sobre la constante de disociación de los tiempos de la PCO_2 . Un valor de pH por sí mismo es difícil de interpretar sin tener información acerca de los valores de bicarbonato y PCO_2 . Del mismo modo, un valor “t” tiene poco sentido a menos que se tenga información acerca de los errores sistemáticos y aleatorios (estadísticos representativos: Sesgo y S_{dif}). Desafortunadamente encontrará a menudo el valor “t” reportado sin ninguna información dada sobre el sesgo o la S_{dif} .

- Valores altos de “t” podrían resultar de un sesgo grande, siendo la S_{dif} pequeña o N grande. Un valor alto de “t” indica que existe un sesgo real, sin embargo, ese sesgo puede ser pequeño y estadísticamente significativo si S_{dif} es pequeño y/o N es grande.
- Valores bajos de “t” podrían resultar de un sesgo pequeño, siendo S_{dif} grande o N pequeño. Un valor bajo de “t” indica que los datos no son suficientes para demostrar que existe una diferencia real, concluyendo así que no existe diferencia. Sin embargo, un sesgo grande podría no ser estadísticamente significativo si la S_{dif} es grande y/o N es pequeño.

La aceptabilidad del desempeño del método depende de si los errores afectarán o no la utilidad clínica de los resultados de examen. La significatividad clínica depende de definir los límites de error permitidos, para luego comparar los errores observados con esos límites. Si los errores observados son menores que los errores permitidos, el desempeño del método es aceptable. Si los errores observados son mayores que los permitidos, el desempeño del método no es aceptable.

¿Cuál es el punto?

¡Necesita tener una estrategia sobre cómo usar la estadística con datos del experimento de comparación de métodos si espera obtener buenos estimados de los errores analíticos de interés! Aquí están las partes importantes de la estrategia del análisis de datos:

- Grafique los datos de comparación en el momento de la recolección para identificar cualquier valor atípico mientras que las muestras de los pacientes aún están disponibles para investigaciones futuras. Puede usar tanto un gráfico de diferencias (diferencia entre los resultados del método evaluado y los resultados del método de comparación en el eje de las ordenadas “y” contra los resultados del método de comparación en el eje de las abscisas “x”) o el gráfico de comparación (resultados del método evaluado como “Y”, resultados del método comparación como “X”).

- Analice los datos de comparación utilizando estadísticas de regresión. Coteje el coeficiente de correlación (idealmente debería ser 0,99 o mayor) para asegurarse de que los cálculos de regresión lineal simple proveen estimados confiables de la pendiente y la intersección. Inspeccione la gráfica de comparación para identificar la presencia de valores atípicos o problemas de no linealidad. Estime el error sistemático a niveles de decisión médica importantes (X_c) calculando los valores Y_c correspondientes, luego tomando las diferencias ($Y_c - X_c$) como la estimación de los errores sistemáticos a cada nivel de decisión médica de interés.
- Si las estadísticas de regresión no son confiables, considere recolectar más resultados de comparación para extender el rango analítico, luego reevalúe el coeficiente de correlación.
- O, analice los datos de comparación con estadísticos “t”. Restrinja la interpretación de resultados a un nivel de decisión médica cercano a la media de los datos.
- O, use cálculos de regresión más sofisticados, tales como la regresión de Deming o la regresión de Passing-Bablok, si están disponibles programas estadísticos especializados.

Recuerde, los exámenes estadísticos pueden proveer estimados de errores sobre los que se pueden hacer una evaluación,, pero no son un sustituto de la evaluación.. La significatividad clínica se determina comparando los estimados estadísticos de los errores al error total permitido definido oportunamente. Una herramienta útil para hacer realizar esta evaluación es la Carta de Decisión de Método que se describirán en detalle en el capítulo 16.

Referencias

1. Westgard JO, Hunt MR. Use and interpretation of statistical tests in method-comparison studies. *Clin Chem* 1973;19:49-57.
2. Stockl D, Dewitte K, Thienpont M. Validity of linear regression in method comparison studies: Is it limited by the statistical model or the quality of the analytical input data? *Clin Chem* 1998;44:2340-6.
3. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. *Clin Chem* 1979;25:432-8.
4. Passing H, Bablock W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. *J Clin Chem Clin Biochem* 1983;21:709-720.
5. Westgard JO. A method evaluation decision chart (MEDx Chart) for judging method performance. *Clin Lab Science* 1995;8:277-83.

13: ¿Cómo examinar fuentes específicas de inexactitud?

Éste capítulo describe experimentos para interferencia y recuperación, que siempre son necesarios si no hay un método comparativo disponible y proveen a menudo información útil junto con los datos obtenidos a partir del experimento de comparación de métodos. Estos experimentos le permiten examinar fuentes específicas que podrían causar inexactitud en el nuevo método. Debido a que estos experimentos no son requeridos para métodos no modificados, podría considerar éste capítulo de lectura opcional.

Objetivos:

- Aprender procedimientos prácticos para validar interferencia y recuperación.
- Identificar los factores importantes para realizar estudios de interferencia y recuperación.
- Distinguir entre procedimientos experimentales de interferencia y recuperación.
- Distinguir entre los cálculos estadísticos apropiados para datos de experimentos de interferencia y recuperación.
- Describir error sistemático constante apropiadamente.
- Describir error sistemático proporcional apropiadamente.

Materiales del capítulo:

- VM – Los experimentos de interferencia y recuperación, por James O. Westgard, PhD
- Problema: Datos de validación de un método de colesterol

Cosas qué hacer:

- Estudiar los materiales.
- Trabajar en el problema de colesterol.

Preguntas de Autoevaluación:

- ¿Cuál es la diferencia entre un experimento de recuperación y uno de interferencia?
- ¿Qué interferencias comunes se estudian usualmente?
- ¿Cuál es la manera adecuada de calcular la recuperación?

Validación de Métodos: Los Experimentos de Interferencia y Recuperación

James O. Westgard, PhD

Los estudios de validación de métodos para exámenes no modificados de complejidad moderada o alta tienden a enfocarse en los experimentos de linealidad o rango reportable, de replicación y de comparación de métodos, los cuales han sido descritos en capítulos anteriores. Sin embargo, nuestro plan experimental recomienda que también se realicen experimentos de interferencia y recuperación para estimar los efectos de materiales específicos en la exactitud o el error sistemático de un método. Estos dos experimentos pueden ser realizados rápidamente para medir fuentes específicas de error. Como tal, complementan las estimaciones del error del experimento de comparación de métodos. Sin embargo, podría saltarse estos materiales, o parte de estos materiales, hasta que aquellos experimentos se vuelvan de interés para sus aplicaciones.

Experimento de Interferencia

Propósito

El experimento de interferencia se realizar para estimar el error sistemático causado por otros materiales que pudieran estar presentes en el espécimen a analizar. Describimos estos errores como errores sistemáticos constantes debido a que una concentración determinada del material interferente causará generalmente una cantidad constante de error, sin importar la concentración del analito a medir en el espécimen que se está examinando. No obstante, se espera que conforme cambie la concentración del material interferente, el tamaño del error constante cambie.

Factores a considerar

El procedimiento experimental está ilustrado en la figura abajo. Un par alícuotas de la muestras de examen se preparan para su análisis por el método que se está estudiando. La primera alícuota se prepara añadiendo una solución del material que se sospecha es interferente (llamado “interferente”, ilustrado por “I” en la figura) a una muestra de un paciente que contenga el analito a medir (ilustrado por “A” en la figura). Una segunda alícuota se prepara diluyendo la alícuota de la muestra original con solvente puro o una solución diluyente que no contenga la interferencia sospechada. Ambas alícuotas se analizan por el método de interés para ver si hay alguna diferencia en los valores debido a la adición de la interferencia sospechada.

Solución del analito.

Se pueden utilizar soluciones estándar, especímenes de pacientes o pools de pacientes. Recomendamos un procedimiento general usando especímenes de pacientes dado que están convenientemente disponibles en un laboratorio clínico y contienen muchas de las sustancias encontradas en un espécimen real.

Replicados.

Es una buena práctica hacer mediciones por duplicado en todas las muestras debido a que el error sistemático se revela por las diferencias entre muestras apareadas. Pequeñas diferencias podrían estar solapadas por el error aleatorio causado por la imprecisión del método. Hacer replicados de las mediciones de los pares de muestras de varios especímenes permite que el error sistemático sea estimado a partir de las diferencias entre los valores promedio, que estarán menos afectados por el error aleatorio del método.

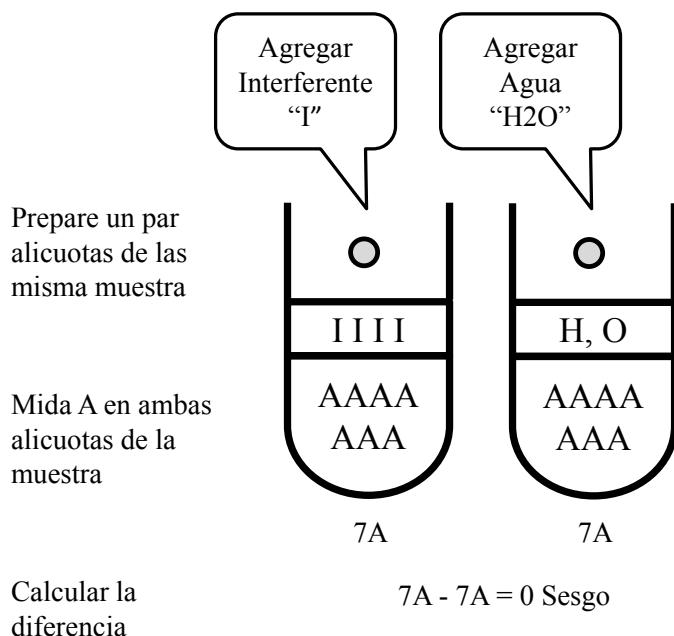
Solución interferente.

Para materiales solubles, es conveniente usar soluciones estándar que sean capaces de introducir la interferencia a concentraciones conocidas. Para algunas interferencias comunes, tales como la lipemia y la hemólisis, se utilizan a menudo muestras o *pools* de pacientes.

Adición del volumen del interferente.

El volumen añadido del interferente debería ser pequeño respecto al volumen de la muestra original para minimizar la dilución del espécimen del paciente. Sin embargo, la *magnitud* de la dilución no es tan importante como mantener *exactamente las mismas diluciones* (volúmenes finales) para el par de las muestras a examinar.

El Experimento de Interferencias



Realización del pipeteo.

La precisión es más importante que la exactitud debido a que es esencial mantener exactamente la misma cantidad de volumen (volumen final) en el par de muestras a examinar.

Concentración del material del interferente.

La cantidad de interferente añadida debería alcanzar un nivel marcadamente elevado, de preferencia cercano de la concentración máxima esperada en la población de pacientes. Por ejemplo, al analizar cómo afecta el ácido ascórbico un método de glucosa, podría ser usada una concentración cercana a 15 mg/dL debido a que representa la concentración máxima esperada [1]. Si se observa un efecto al nivel máximo, entonces también podría ser de interés examinar concentraciones más bajas y determinar el nivel en el cual la interferencia invalida por primera vez la utilidad de los resultados analíticos.

Interferencias a medir.

Las sustancias a ser medidas son seleccionadas a partir de las especificaciones de desempeño publicadas por los fabricantes en sus insertos y/o documentos similares, información publicada en literatura, los artículos de revisión de materiales interferentes y datos de tabulados o bases de datos, tales como la tabla extensiva compilada por Young y cols., que también contiene una bibliografía comprensiva.

También es una buena práctica examinar interferencias comunes tales como la bilirrubina, la hemólisis, la lipemia y los aditivos (conservantes) y anticoagulantes utilizados en la recolección del espécimen.

- La bilirrubina puede ser medida por adición de una solución estándar de bilirrubina.
- La hemólisis se mide a menudo removiendo una alícuota de una muestra, luego hemolizándola mecánicamente, o congelando, y descongelando el espécimen antes de remover una segunda alícuota.
- La lipemia se puede medir añadiendo una emulsión de grasa comercial o analizando los especímenes de pacientes con lipemia antes y después de un proceso de ultra centrifugación [3, ver procedimiento recomendado por el CLSI].
- Los aditivos para los tubos de recolección de especímenes pueden ser estudiados convenientemente tomando una muestra de sangre total, luego dispensándola en alícuotas en una serie de tubos que contengan diferentes aditivos.

Método comparativo.

Recomendamos que las muestras de interferencia también sean analizadas por el método comparativo, particularmente cuando el método comparativo es un método de que se trabaja de rutina. Si ambos métodos sufren efectos provocados por la misma interferencia, ésta interferencia no puede ser motivo suficiente para rechazar el método. El método de examen podría tener otras características que mejorarían aún el desempeño general del examen. Si la razón para cambiar los métodos es librarse de una interferencia, luego, por supuesto, los datos de interferencia deberían ser usados para rechazar el nuevo método.

Cálculos de los datos

El análisis de datos es equivalente a los cálculos del “estadístico t para datos apareados” en un estudio de comparación de métodos y puede ser llevado a cabo con el mismo programa estadístico. Sin embargo, el número de muestras apareadas será mucho menor que los 40 especímenes requeridos típicamente en el estudio de comparación de métodos. Note también que las estadísticas de regresión no son apropiadas aquí dado que los datos probablemente no demostrarán un rango analítico amplio. Aquí hay un procedimiento paso a paso para calcular los datos:

1. Tabule los resultados para las muestras apareadas.
 - Muestra A: “T” añadida = 110, 112 mg/dL; dilución = 98, 102 mg/dL;
 - Muestra B: “T” añadida = 106, 108 mg/dL; dilución = 93, 95 mg/dL;
 - Muestra C: “T” añadida = 94, 98 mg/dL; dilución = 80, 84 mg/dL.
2. Calcular el promedio de los replicados.
 - Muestra A: “T” añadida = 111 mg/dL; dilución = 100 mg/dL;
 - Muestra B: “T” añadida = 107 mg/dL; dilución = 94 mg/dL;
 - Muestra C: “T” añadida = 96 mg/dL; dilución = 82 mg/dL.

3. Calcule las diferencias entre los resultados de las muestras pareadas.

- Muestra A diferencia = 11 mg/dL;
- Muestra B diferencia = 13 mg/dL;
- Muestra C diferencia = 14 mg/dL.

4. Calcule la diferencia promedio de todos los especímenes examinados a una concentración o nivel de interferencia dados.

- Interferencia promedio = 12.7 mg/dL.

Criterios para un desempeño aceptable

La evaluación de la aceptabilidad se hace comparando el error sistemático observado con la cantidad de error permitido para el examen. Por ejemplo, se supone que un examen de glucosa es correcto dentro de un margen de error del 10% de acuerdo a los criterios de CLIA para desempeño aceptable. En el extremo superior del rango de referencia (110 mg/dL), el error permitido sería 11.0 mg/dL. Dado que la interferencia observada de 12.7 mg/dL es mayor que el error permitido, el desempeño de éste método no es aceptable.

Experimento de Recuperación

Los estudios de recuperación son técnicas clásicas para validar el desempeño de cualquier método analítico. Sin embargo, su uso en laboratorios clínicos ha sido cargado de problemas debidos a desempeños inapropiados del experimento, cálculos inadecuados de los datos e interpretación inadecuada de los resultados. Los estudios de recuperación, por lo tanto, son utilizados selectivamente y no tienen alta prioridad cuando está disponible otro método analítico para efectuar un experimento de comparación. Sin embargo, podrían ser todavía útiles para ayudar a entender la naturaleza de cualquier sesgo revelado en el experimento de comparación de métodos. En ausencia de un método de comparación confiable, los estudios de recuperación deberían tomar más importancia.

Propósito

El experimento de recuperación se lleva a cabo para estimar el error sistemático proporcional. Éste es el tipo de error cuya magnitud se incrementa conforme aumenta la concentración del analito. El error es causado a menudo por una sustancia en la matriz de la muestra que reacciona de manera similar al analito

a medir y por tanto compite con el reactivo analítico. El experimento también podría ser de ayuda para investigar soluciones de calibración cuyos valores asignados son usados para establecer los puntos de ajuste del instrumento (ajuste de calibración).

Factores a considerar

El procedimiento experimental se resume en la figura a la derecha. Note que las muestras apareadas están preparadas en forma similar al experimento de interferencia. La diferencia importante es que la solución añadida contiene el analito "A" a medir (mostrado como A) en lugar del material interferente (mostrado como "I" en la figura anterior). La solución añadida es a menudo un estándar o solución de calibración del analito a medir. Ambas alicuotas son analizadas entonces por el método de interés.

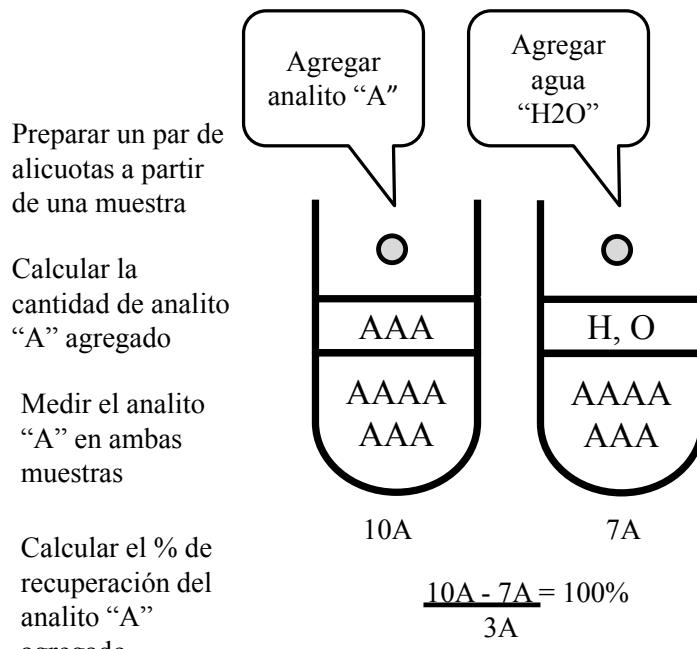
Volumen añadido del estándar.

Es importante mantener pequeño el volumen del estándar en relación al volumen del espécimen original del paciente para minimizar la dilución de la matriz del espécimen original. De otro modo, el error podría cambiar conforme se diluya la matriz. Recomendamos que la dilución del espécimen original no sea mayor a 10%. Para un procedimiento práctico, añada 0.1 ml de solución estándar a 0.9 ml o 1.0 ml del espécimen del paciente.

Exactitud del pipeteo.

Ésta es crítica debido a que la concentración del analito añadido será calculada a partir del volumen del estándar y del volumen espécimen original del paciente. El trabajo experimental debe llevarse a cabo cuidadosamente. Se deberían emplear pipetas de alta calidad. Debe prestarse también particular atención a su limpieza, llenado y tiempo de descarga.

El Experimento de Recuperación



Concentración del analito añadido.

Una guía práctica es añadir una cantidad suficiente del analito a medir como para alcanzar el siguiente nivel de decisión médica del examen. Por ejemplo, para especímenes de glucosa con valores de referencia normales en el rango de 70 a 110 mg/dL, una adición de 50 mg/dL elevaría las concentraciones a un rango de 120 a 160 mg/dL, que están en el rango más alto donde la interpretación médica de los exámenes de glucosa será crítica. También es importante considerar la variabilidad (precisión) del método. Un nivel pequeño de adición estará más afectado por la imprecisión del método que un nivel grande.

Concentración de la solución estándar.

Dada la importancia de añadir un pequeño volumen para minimizar el efecto de la dilución, será deseable usar soluciones estándar con concentradas. Para nuestro ejemplo de glucosa, una solución estándar con 500 mg/dL sería necesaria para hacer una adición de 50 mg/dL, asumiendo que se añaden 0.1 ml del estándar a 0.9 ml de un espécimen de un paciente. Una solución estándar de 1000 mg/dL sería necesaria para hacer una adición de 100 mg/dL. La concentración de la solución estándar puede ser calculada una vez que los volúmenes de la adición del estándar y del espécimen del paciente han sido decididos. Si se adopta un procedimiento general de 0.1 ml de estándar y 0.9 ml de espécimen, entonces la concentración de la solución estándar necesitará ser 10 veces el nivel deseado de adición.

Número de replicados de medición por espécimen.

Se debería efectuar mediciones por replicados sobre todas las muestras porque el error aleatorio propio de las mediciones con frecuencia dificulta evidenciar la presencia de errores sistemáticos pequeños. Como regla general, lleve a cabo mediciones por duplicado. Si la adición del estándar es baja en relación a la concentración de los espécímenes originales, sería aconsejable realizar mediciones por triplicado o cuadruplicado.

Número de espécímenes de pacientes examinados.

Esto depende de la reacción competitiva que podría causar un error sistemático. Por ejemplo, si el objetivo es determinar si proteínas séricas afectan la reacción analítica, entonces sólo se requieren investigar unos pocos espécímenes de pacientes dado que todos contienen proteínas. Si el objetivo es determinar si los metabolitos de algún fármaco afectan la recuperación, entonces se deben examinar varias muestras de pacientes diferentes.

Verificación de la técnica experimental.

Es una buena práctica analizar las muestras de recuperación por los métodos de examen y de comparación. Hay problemas ocasionales causados por la inestabilidad de las soluciones estándar, errores en la preparación de las muestras, intercambio de muestras y errores en el cálculo de los datos. Si el método de comparación muestra la misma recuperación que el método de examen, los resultados de éste experimento son de valor limitado al evaluar la aceptabilidad del método de examen.

Cálculo de datos

La recuperación debería ser expresada como un porcentaje ya que el objetivo experimental es estimar error sistemático proporcional, que es un tipo de error porcentual. La recuperación ideal es del 100%. La diferencia entre 100 y la recuperación observada (en porcentaje) es el error sistemático proporcional. Por ejemplo, una recuperación del 95% corresponde a un error proporcional del 5%.

Los cálculos de recuperación son difíciles y a menudo se realizan incorrectamente, incluso en estudios publicados en revistas científicas. Aquí hay un procedimiento paso a paso para calcular los datos:

1. Calcule la cantidad de analito añadido multiplicando la concentración de la solución estándar por el factor de dilución (ml estándar)/(ml estándar + ml espécimen).

- Por ejemplo, para un método de calcio, si 0.1 ml de un estándar de 20 mg/dL se añade a 1.0 ml de suero, la cantidad añadida es $20 * (0.1 / 1.1)$ ó 1.82 mg/dL.

2. Promedie los resultados para los replicados de medición de cada muestra.

- Muestra A adición = $(11.4 / 11.6) / 2 = 11.5$ mg/dL;
- Muestra A dilución = $(9.7 + 9.9) / 2 = 9.8$ mg/dL;
- Muestra B adición = $(11.2 + 11.0) / 2 = 11.1$ mg/dL;
- Muestra B dilución = $(9.5 + 9.5) / 2 = 9.5$ mg/dL.

3. tome la diferencia entre la muestra con adición y la muestra con dilución.
 - Muestra A adición = 11.5, Muestra A dilución = 9.8, diferencia = 1.7 mg/dL;
 - Muestra B adición = 11.1, Muestra B dilución = 9.5, diferencia = 1.6 mg/dL;
4. Calcule la recuperación para cada espécimen como la “diferencia” [paso 3] dividida por la cantidad añadida [paso 1] en porcentaje.
 - $(1.7 \text{ mg/dL}/1.82 \text{ mg/dL})100 = 93.4\%$ de recuperación;
 - $(1.6 \text{ mg/dL}/1.82 \text{ mg/dL})100 = 87.9\%$ de recuperación;

[Note la variabilidad de estos estimados, que probablemente se deba a la imprecisión del método; podría ser deseable realizar más replicados de la medición o preparar más muestras].

5. Promedie las recuperaciones de todos los especímenes examinados.
 - $(93.4 + 87.9)/2 = 90.6\%$ de recuperación promedio.
6. Calcule el error proporcional.
 - $100 - 90.6 = 9.4\%$ de error proporcional.

Criterios para un desempeño aceptable

El error observado es comparado con la cantidad de error total permitido para el examen. Para calcio, por ejemplo, el criterio de CLIA para un desempeño aceptable es de 1 mg/dL. En el medio del rango de referencia, alrededor de 10 mg/dL, el error total permitido es del 10%. Dado que el error proporcional observado es de 9.4%, el desempeño de nuestro método cumple de manera ajustada con los criterios de CLIA para aceptabilidad.

Resumen sobre interferencia y recuperación

Los experimentos de interferencia y recuperación pueden ser utilizados para valorar los errores sistemáticos de un método. Complementan el experimento de comparación de métodos permitiendo estimados iniciales rápidos de errores específicos — el experimento de interferencia para error sistemático constante y el experimento de recuperación para error sistemático proporcional. En ausencia de un método de comparación, proveen una forma alternativa para estimar

errores sistemáticos.

Las técnicas experimentales son similares, pero el material añadido es diferente. Un material que se sospeche interferente es añadido en el experimento de interferencia, mientras que el analito a medir se añade en el experimento de recuperación.

El cálculo de los datos es diferente. El sesgo entre las muestras apareadas debería ser calculado para los datos de interferencia, de forma similar al cálculo de los estadísticos “t” en el experimento de comparación de métodos. La recuperación promedio porcentual debería ser calculada del experimento de recuperación, siendo cuidadosos al dividir la diferencia entre las muestras apareadas por la cantidad añadida, no por el total después de la adición. El error sistemático proporcional es la diferencia entre 100% y el porcentaje de recuperación observado.

Los experimentos de interferencia son generalmente útiles para evaluar los efectos de condiciones comunes de las muestras, como ictericia, hemólisis, lipemia y aditivos de los tubos de recolección de muestra. Los resultados en la literatura son generalmente confiables.

Los estudios de recuperación se llevan a cabo con menor frecuencia y sus resultados en la literatura son difíciles de interpretar debido a la falta de un procedimiento estandarizado para el cálculo de los datos. Se requiere de mucho cuidado y atención cuando se realizan estudios de recuperación y también cuando se interpretan sus resultados.

Cuando se hacen juicios sobre el desempeño de un método, los errores observados deberían ser comparados con el error permitido definido oportunamente. La estimación del sesgo a partir de un experimento de interferencia puede ser comparada directamente a un requisito de la calidad analítico expresado en unidades de concentración. La recuperación promedio necesita ser convertida a error proporcional ($100 - \%$ de recuperación) y luego comparada con un requisito de la calidad analítico expresado en porcentaje.

Referencias

1. Katz SM, DiSalvio TV. Ascorbic acid effects on serum glucose values. *JAMA* 1973;224:628.
2. Young DS. Effects of preanalytical variables on clinical laboratory tests. Washington DC:AACC Press, 1993.
3. CLSI Document EP7-A. Interference testing in clinical chemistry. Approved Guideline. CLSI, 940 West Valley Road, Wayne, PA, 2002.

Problema – Datos de Validación de un Método de Colesterol: Recuperación (usada para estimar el error sistemático proporcional)

Se utilizaron seis especímenes de pacientes diferentes (#1-6). Se prepararon dos muestras para examen (base y suplementada con analito). Cada muestra fue medida 4 veces. Las muestras base fueron preparadas diluyendo 0.9 mL de cada espécimen de paciente con 0.1 ml de solución salina. Las muestras con el analito fueron preparadas mezclando 0.9 mL de cada espécimen de paciente con 0.1 mL de un estándar de colesterol de 500 mg/dL.

Calcule la cantidad de colesterol añadida a las muestras suplementadas con analito, la cantidad recuperada de cada muestra y el promedio de recuperación de los 6 especímenes. Recuerde, la recuperación se expresa como %. (Respuestas en la p.306).

Muestras de pacientes	Muestra Base			
	0.9 mL muestra + 0.1 mL salina			
	Resultado 1	Resultado 2	Resultado 3	Resultado 4
1	149	151	153	146
2	180	186	178	187
3	201	204	196	206
4	180	204	184	188
5	160	157	166	159
6	187	182	191	201

Muestras de pacientes	Muestra Suplementada			
	0.9 mL muestra + 0.1 mL Estándar de Colesterol			
	Resultado 1	Resultado 2	Resultado 3	Resultado 4
1	204	196	208	194
2	224	222	228	240
3	255	243	257	258
4	235	246	233	248
5	206	207	210	208
6	235	242	246	246

Nota: Los resultados en la tabla están expresados en mg/dL

Problema – Datos de Validación de un Método de Colesterol: Interferencia (usada para estimar el error sistemático constante)

Se utilizaron seis especímenes de pacientes diferentes (#1-6). Se prepararon dos muestras para examen (Base y Suplementada con interferente). Cada muestra fue medida 4 veces. Las muestras base fueron preparadas diluyendo 0.9 mL de cada espécimen de paciente con 0.1 ml de solución salina. Las muestras suplementadas con interferente fueron preparadas mezclando 0.9 mL de cada espécimen de paciente con 0.1 mL de un estándar de bilirrubina de 100 mg/dL.

Calcule la cantidad de interferencia (sesgo) para cada muestra y el sesgo promedio para todas las muestras. Recuerde, la interferencia se expresa en unidades de concentración (mg/dL). (Respuestas en la p.306).

Muestras de pacientes	Muestra Base			
	0.9 mL muestra + 0.1 mL salina			
	Resultado 1	Resultado 2	Resultado 3	Resultado 4
1	206	213	223	215
2	220	228	223	210
3	299	287	297	297
4	169	171	167	178
5	250	248	257	252
6	227	221	224	230

Muestras de pacientes	Muestra Suplementada			
	0.9 mL muestra + 0.1 mL Estándar de Bilirrubina			
	Resultado 1	Resultado 2	Resultado 3	Resultado 4
1	221	222	230	229
2	233	241	228	237
3	306	304	302	296
4	186	184	181	183
5	242	265	271	262
6	236	229	237	242

Nota: Los resultados en la tabla están expresados en mg/dL

14: ¿Cuál es el valor más bajo confiable para un método?

El experimento de rango reportable es usado para caracterizar qué tan alto puede ser un valor de examen confiable, esto es, dentro del rango reportable. El extremo más bajo está caracterizado por un valor de cero. Con ciertos exámenes, tales como drogas de abuso, TDMs, y algunos marcadores cardiacos, es necesario determinar más exactamente el valor mínimo que puede ser medido, esto es, el límite de detección del método. Esto se logra realizando un experimento de límite de detección.

Objetivos:

- Aprender un procedimiento práctico para determinar el límite de detección de un método.
- Distinguir entre los conceptos Límite de Blanco (LB), Límite de Detección (LD), Sensibilidad Funcional (SF) y Límite de Cuantificación (LC).
- Distinguir los procedimientos experimentales para éstas características diferentes.

Materiales del capítulo:

- VM – El experimento de límite de detección,
por James O. Westgard, PhD
- Problema: Datos de validación de un método de colesterol

Cosas qué hacer:

- Estudiar los materiales.
- Trabajar con los datos provistos para el problema de Colesterol.

Preguntas de Autoevaluación:

- ¿Cuándo usted necesita realizar un experimento de límite de detección?
- ¿Qué estimación del límite de detección dará generalmente el valor más alto?
- ¿Por qué el límite de detección (LD) siempre es más alto que el límite de blanco (LB)?

Validación de Métodos: El Experimento de Límite de Detección

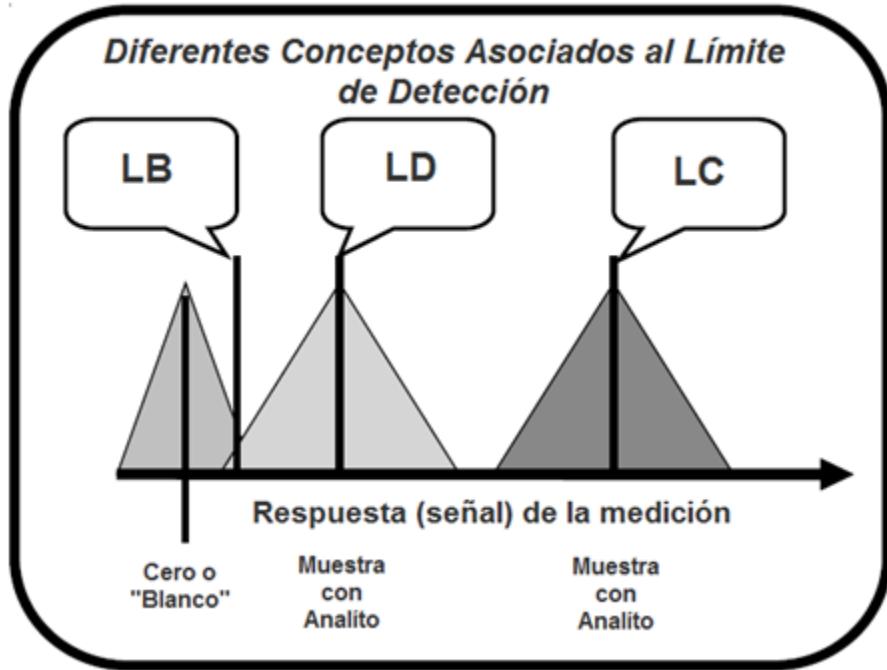
James O. Westgard, PhD

Propósito

El experimento de límite de detección está previsto para estimar la menor concentración a la que puede ser medido un analito. Éste límite inferior de concentración es obviamente de interés en exámenes forenses de drogas, donde la presencia o ausencia de una droga puede ser información crítica del examen. El desempeño analítico a bajas concentraciones también es importante para marcadores tumorales, tales como el antígeno prostático específico (PSA, por sus siglas en Inglés), donde el monitoreo de los valores de los pacientes después del tratamiento podría ser útil para valorar un “relapso bioquímico” [1].

Las regulaciones de laboratorios en E.U.A. requieren que el límite de detección (o sensibilidad analítica) sea establecido sólo para métodos *non-waived* que han sido modificados por el laboratorio y sistemas analíticos que no han sido sujetos a la aprobación de la FDA, tales como métodos desarrollados en el propio laboratorio. Las buenas prácticas de laboratorio dictan que el límite de detección debe ser verificado, cuando sea relevante, por ejemplo, todos los exámenes de drogas terapéuticas y exámenes; TSH e inmunoensayos similares; marcadores cardíacos tales como las troponinas; PSA y otros marcadores de cáncer. El límite de detección no es importante para exámenes tales como la glucosa, el colesterol, enzimas y otros constituyentes donde hay rangos “normales” o de referencia que son relevantes para la interpretación de resultados de exámenes.

¡La terminología en ésta área es un desastre! Al establecer sus especificaciones de desempeño, los fabricantes a menudo usan una variedad de términos, tales como sensibilidad, sensibilidad analítica, límite mínimo de detección, límite inferior de detección, límite de blanco, límite biológico de detección, límite de detección, sensibilidad funcional y límite de cuantificación [2-6]. La guía de la CLSI EP17-A recomienda estandarizar la terminología con el uso de los siguientes tres parámetros:



- **Límite de Blanco (LB):** El resultado más alto de una medición que probablemente se observará (con una probabilidad dada) para una muestra de blanco; estimado típicamente como un límite de confianza del 95 %, (distribución normal, un lado), por el valor medio del blanco más 1.65 veces la desviación estándar del blanco.
- **Límite de Detección (LD):** La cantidad de analito más baja en una muestra que puede ser detectada con una cierta probabilidad, aunque tal vez no cuantificada como un valor exacto; estimada como un límite de confianza del 95 % (distribución normal, un lado), por el valor medio del blanco más 1.65 veces la desviación estándar de una muestra de concentración baja.
- **Límite de Cuantificación (LC)/Límite inferior de cuantificación:** La cantidad más baja de analito que puede ser determinada cuantitativamente con precisión y veracidad aceptables, bajo condiciones experimentales dadas; la concentración de analito a la cual el 95% del límite del error total, esto es el sesgo más dos veces la desviación estándar, cumple con la meta requerida o dada para el error permitido.

Note que el LB reemplaza el término anterior Límite Inferior de Detección (LID) y es estimado de la manera similar. La principal diferencia es que el múltiplo por el cual la desviación estándar es multiplicada es 1.65, en lugar de 2 ó 3 que se usa comúnmente con LID.

Del mismo modo, el LD reemplaza el Límite Biológico de Detección (LBD), siendo nuevamente la principal diferencia el múltiplo empleado para multiplicar el valor del desvío estándar que es de 1.65, en lugar del de 2 ó 3 usado comúnmente con LBD.

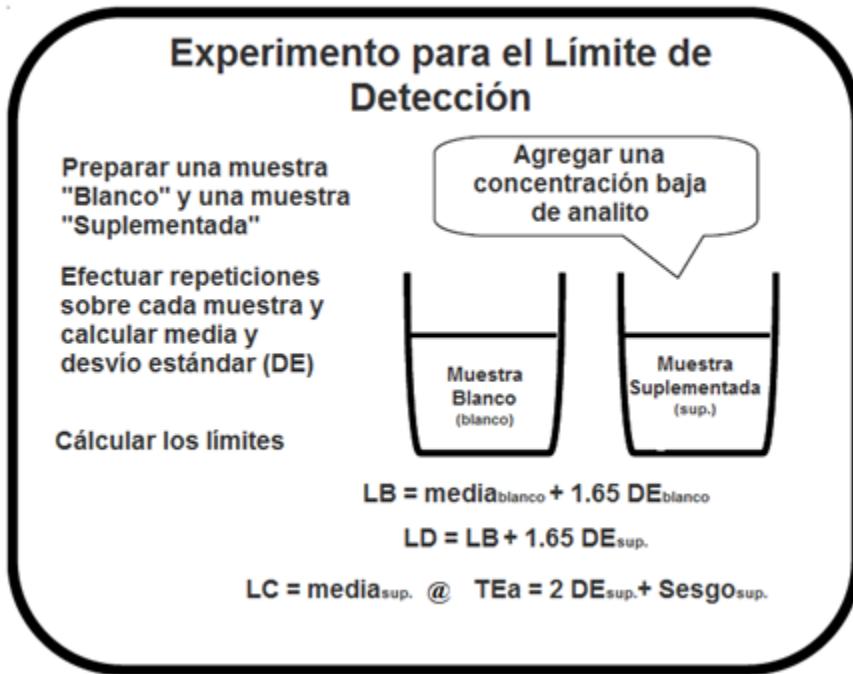
El LC es de cierto modo similar al concepto anterior de Sensibilidad Funcional:

Sensibilidad Funcional (SF): La concentración del analito a la que el CV del método es 20% [2].

Ambos representan las concentraciones de analito a las que se alcanza cierto nivel de desempeño analítico. Para Sensibilidad Funcional, ese nivel es la concentración a la que se alcanza un CV del 20%. Para el LC, ese nivel es la concentración del analito a la que la suma del sesgo del método más dos veces la desviación estándar del método satisface una meta de error permitido. Es más complicado estimar el LC porque es más difícil estimar el sesgo del método. También es improbable que muchos fabricantes establezcan sus especificaciones para LC, porque no les es requerido por la FDA volcar esas especificaciones en el etiquetado (inserto) juntos con las demás especificaciones de desempeño que definen la calidad que debería ser alcanzado por un examen.

Factores a considerar

La figura de abajo provee una descripción general del procedimiento experimental. Se analizan generalmente dos tipos diferentes de muestras. Una muestra es un “blanco” con una concentración cero del analito de interés. La segunda es una muestra “suplementada” que tiene una concentración baja del analito de interés. En algunas situaciones, particularmente la estimación de la SF o del LC, muchas muestras de diferentes concentraciones podrían ser necesarias a concentraciones progresivamente mayores del analito. Tanto el blanco como las muestras con diferentes concentraciones del analito de interés se miden repetidamente en un tipo de experimento de replicación, luego se calculan las medias y los desvíos estándar de los valores observados, y la estimación del límite de detección realiza a partir de esos resultados.



Solución blanco. Una alícuota de la solución blanco se usa generalmente para el “blanco” y otra alícuota se usa para preparar una muestra suplementada. Idealmente, la solución “blanco” debería tener la misma matriz que las muestras de pacientes regulares. Sin embargo, también es común usar el “estándar cero” de una serie de calibradores como el blanco y el estándar más bajo como la muestra “suplementada”.

Muestra Suplementada. Al verificar la especificación establecida por el fabricante para el límite de detección de un método, la cantidad de analito añadida a la solución “blanco” debería representar o estar próximo a la declarada por el fabricante en sus especificaciones. Para establecer el límite de detección, podría ser necesario preparar varias muestras suplementadas cuyas concentraciones estén en el rango analítico del límite de detección esperado. Para algunos exámenes, podría ser de interés usar muestras de pacientes que estén libres de enfermedad bajo tratamiento (por ejemplo, sueros de EPA de pacientes tratados para cáncer de próstata) [4].

Número de replicados de las mediciones. No hay una guía contundente y rápida, pero se recomiendan a menudo 20 replicados de mediciones en la literatura. Éste número es razonable dado que el experimento de límite de detección es un caso especial del experimento de replicación, donde 20 mediciones se aceptan generalmente como el mínimo. Los lineamientos de

la guía de la CLSI también sugieren que el laboratorio haga 20 replicados para verificar una especificación de un fabricante, pero recomienda que los fabricantes trabajen con un mínimo de 60 replicados para establecer una especificación.

El periodo de tiempo del estudio. Un estudio en condiciones de repetibilidad (intra corrida) o a corto plazo se lleva a cabo usualmente cuando el enfoque principal es el desempeño del método en una solución blanco. Un periodo de tiempo más largo, que represente un desempeño en condiciones de precisión intermedia (entre días), se recomienda cuando el enfoque principal está en la muestra suplementada [2]. Los lineamientos de la guía de la CLSI [7] recomiendan que el LD sea estimado a partir de los datos obtenidos en un periodo de “varios días” y el LC de datos de al menos 5 corridas, asumido como un periodo mayor a 5 días. Por lo tanto se deben hacer mediciones múltiples a diario durante cinco días.

Cantidad a ser estimada. Para los laboratorios, será generalmente más práctico verificar o establecer el LB o, preferentemente, el LD. En casos en los que el laboratorio quiera verificar especificaciones para SF o LC, será mejor seguir el protocolo experimental descrito en el reporte del fabricante o en un estudio científico.

Ejemplo de estimaciones

Para ilustrar los diferentes conceptos de límites de detección (incluyendo la SF), considere un caso donde 20 replicados provean una S de 1.0 $\mu\text{g}/\text{L}$ para un blanco y también para bajas concentraciones de muestras hasta 10 $\mu\text{g}/\text{L}$. Asuma que éste es un examen de fármacos, donde el criterio de CLIA para desempeño aceptable es 25%.

LB será igual a $1.65 * 1.0 \mu\text{g}/\text{L}$ ó $1.65 \mu\text{g}/\text{L}$

LD será igual a LB más $1.65 * 1.0 \mu\text{g}/\text{L}$, ó sea $3.30 \mu\text{g}/\text{L}$.

La SF será igual a la concentración donde el CV es 20%, el cual puede ser calculado de la ecuación para un CV, por ejemplo, $\text{CV} = 20\% = S * 100 / \text{SF}$ igual a $1.0 \mu\text{g}/\text{L} * (100\% / 20\%)$ ó $5.0 \mu\text{g}/\text{L}$.

Para el LC, asuma que el sesgo es 0.0, así el TEa = $25\% = 2 * \text{CV}$ y un CV de 12.5% ($25\% / 2$) será observado a una concentración de $S * (100\% / 12.5\%)$ ó $1.0 * 8$ ó $8.0 \mu\text{g}/\text{L}$.

En la práctica, es necesario obtener estimados experimentales de la S para varias concentraciones de la muestra, por ej., 0.0 µg/L, 3.0 µg/L, 6.0 µg/L y 9.0 µg/L. Es improbable que la S sea exactamente la misma, así que el LB será estimado a partir del A de la muestra de concentración 0.0 µg/L, el LD usando la S de la muestra de concentración 3.0 µg/L, la SF de la muestra de 6.0 µg/L y el LC de la muestra de 9.0 µg/L. Un acercamiento para estimar la SF y el LC es crear un “perfil de precisión”, esto es, graficar el CV en función de las concentraciones de las distintas muestras o diluciones, dibujar una curva a lo largo de los puntos, luego interpolar en esa curva para identificar la concentración que corresponda a un CV de 20% (SF) y luego la concentración donde el CV es 12.5% (LC).

Verificación del reporte del LC y el LD

Los lineamientos de la guía de la CLSI recomiendan que los laboratorios verifiquen las especificaciones provistas por los fabricantes haciendo 20 mediciones. Para el LB, la especificación se verifica si 3 o menos de los 20 resultados del blanco excedieron el LB declarado por el fabricante. Para el LD, la especificación provista por el fabricante se verifica si no más de 1 de los 20 replicados de la muestra suplementada (con la concentración del LD o próxima) es menor al LB. Para emplear un número diferente de replicados de medición, consulte los lineamientos de la guía de la CLSI [7].

Resumen de Comentarios

Para validar las especificaciones declaradas por el fabricante para el límite de detección, es importante reconocer de qué formas fue estimada la especificación, los datos necesarios para verificarla y el manejo apropiado de los mismos para una correcta verificación. Muchos fabricantes escogen el LB dado que es el más sencillo de estimar y da el número más bajo — una ventaja de mercadeo. Para aplicaciones clínicas, generalmente son más útiles las estimaciones del LD o el LC. Más fabricantes establecerán más especificaciones para LD que para LC; así que tal vez el LD es la estimación más práctica. La guía EP17-A de la CLSI brinda lineamientos definitivos para estos conceptos, los procedimientos experimentales y el análisis de datos, así que la mayoría de los fabricantes seguirán esos lineamientos.

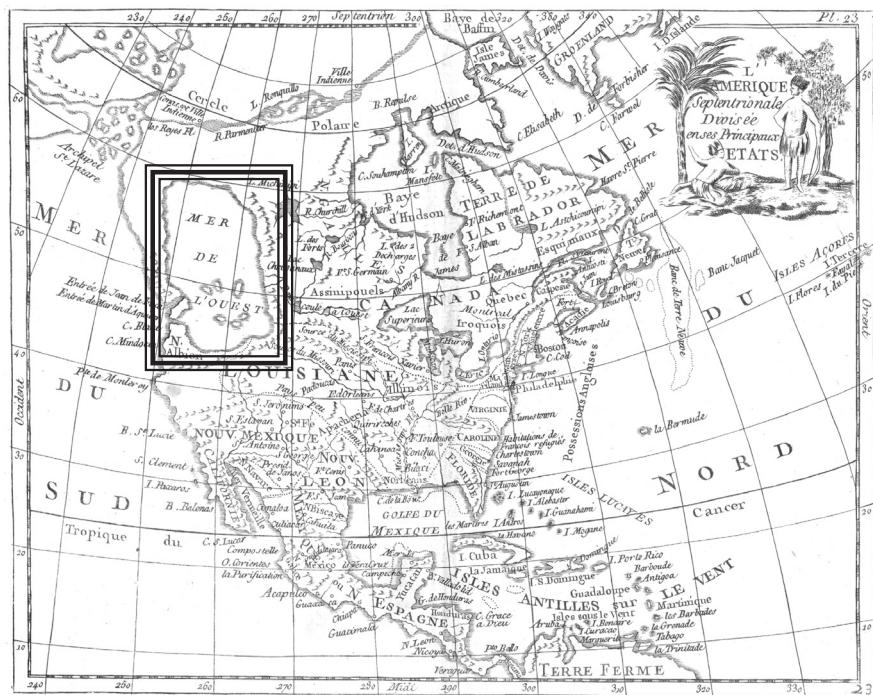
Referencias

1. Diamandis EP, Yu J, Melegos DN. Ultrasensitive prostate-specific antigen assays and their clinical application. Clin Chem 1996;42:853-857.
2. Stamey TA. Lower limits of detection, biological detection limits, functional sensitivity, or residual cancer detection? Sensitivity reports on prostate-specific antigen assays mislead clinicians. Clin Chem 1996;42:849-852.
3. Lawson GM. Defining limit of detection and limit of quantitation as applied to drug of abuse testing: striving for a consensus. Clin Chem 1994;40:1218-1219.
4. Armbruster DA, Tillman MD, Hubbs LM. Limit of detection (LOD)/limit of quantitation (LOQ): Comparison of the empirical and the statistical methods exemplified with GC-MS assays of abused drugs. Clin Chem 1994;40:1233-1238.
5. Klee GG, Dodge LA, Zincke H., Oesterling JE. Measurement of serum prostate specific antigen using Imx prostate specific antigen assay. J Urol 1994;151:94-98.
6. Spencer CA, Takeuchi M, Kazarosyan M, MacKenzie F, Beckett GH, Wilkinson E. Interlaboratory/Intermethod differences in functional sensitivity of immunometric assays of thyrotropin (TSH) and impact on reliability of measurement of subnormal concentrations of TSH. Clin Chem 1995;41:367-374.
7. CLSI EP17-A. Protocols for Determination of Limits of Detection and Limits of Quantitation. Clinical Laboratory Standards Institute, Wayne PA, 2003.

Problema: Datos de Validación de un Método de Colesterol: Límite de Detección

Apartir de los datos de abajo calcule el límite de blanco (LB), el límite de detección (LD), y la sensibilidad funcional (SF). (Respuestas, pág. 306).

Muestra 0	Muestra 2	Muestra 4	Muestra 6	Muestra 8
-0.23	2.84	5.51	6.46	6.92
0.12	2.26	5.07	5.14	6.29
-0.40	1.25	3.15	3.78	8.67
0.13	3.52	2.72	7.28	7.86
0.05	1.95	3.67	6.20	5.24
0.18	2.71	5.35	5.00	7.38
-0.41	3.07	5.22	5.54	7.41
0.20	2.47	4.46	6.44	6.06
-0.32	1.95	2.55	6.71	7.53
-0.08	1.94	3.61	7.17	10.11
-0.41	1.93	3.79	4.92	8.70
-0.39	2.45	3.21	4.40	6.37
-0.32	2.38	3.42	6.11	7.54
-0.37	1.37	4.73	7.51	9.30
-0.17	1.58	4.07	6.38	8.10
0.38	2.46	3.47	5.55	7.87
-0.31	2.55	3.28	8.19	8.20
-0.41	2.57	5.67	8.48	7.99
0.10	2.27	4.99	5.45	9.38
-0.30	1.57	4.42	7.61	8.22



Por último, pero no menos importante!

L'AMERIQUE SEPTENTRIONALE, by De la Porte, Paris, 1786.

El noroeste del Pacífico fue una de las últimas zonas por explorar y dio lugar a algunos de los mitos más grandes, como el gran “Mar del Oeste” que aparece en estos mapas. Es probable que los experimentos de interferencia, recuperación, y el de límite de detección sean también los últimos que usted va a realizar usted va a realizar. Sin embargo, son críticos para ciertos métodos en su laboratorio. Recuerde que son discutidos en estos dos últimos capítulos; regrese y repase este material cuando sean relevantes para sus propias aplicaciones.

15: ¿Cómo se verifica el intervalo de referencia de un método?

Una última consideración es verificar que el intervalo de referencia es apropiado para la población de pacientes. En éste capítulo, Patricia L. Barry, BS, MT (ASCP), presenta un procedimiento práctico para evaluar la validez de las recomendaciones del fabricante para intervalos de referencia.

Objetivos:

- Describir 4 modelos para la verificación o transferencia de intervalos de referencia.
- Identificar las condiciones de aplicación de los diferentes modelos.
- Comparar los requisitos de datos para los diferentes modelos.
- Comparar los cálculos a efectuar a partir de los datos y su presentación.
- Organizar los acercamientos en un plan para su aplicación práctica.

Materiales del capítulo:

- VM — Transferencia del intervalo de referencia, por Patricia L. Barry, BS, MT(ASCP) y James O. Westgard, PhD

Cosas qué hacer:

- Estudiar los materiales.

Preguntas de Autoevaluación:

- ¿Cuántos pacientes se necesitan por lo general para verificar un intervalo de referencia?
- ¿Cuál es el número mínimo de especímenes de pacientes para establecer un intervalo de referencia?
- ¿Cómo se puede transferir un intervalo de referencia desde un método de comparación?

Validación de Métodos: Transferencia del Intervalo de Referencia

Patricia L. Barry, BS, MT (ASCP), y James O. Westgard, PhD

Propósito

El intervalo de referencia es la última característica a ser estudiada en el proceso de validación de métodos. Se estudia en general al final porque el intervalo de referencia en sí no entra en la decisión de la aceptabilidad del método y el estudio no es necesario cuando el desempeño del método es inaceptable. Si el desempeño del método es aceptable, entonces es importante valorar el(os) intervalo(s) de referencia para fortalecer a la interpretación de los resultados de pacientes.

Trasfondo

Un intervalo de referencia es establecido por lo general ensayando especímenes que son obtenidos de individuos que cumplen con criterios de selección que han sido definidos cuidadosamente (muestra de la población de referencia). Protocolos tales como los del Grupo de Expertos en Teoría de los Valores de Referencia de la *International Federation of Clinical Chemistry* (IFCC) [1-6] y los del *Clinical Laboratory Standards Institute* (CLSI)[7] delinean procesos lógicos y sistemáticos que emplean una muestra de la población de referencia seleccionada cuidadosamente para establecer intervalos de referencia. Estos protocolos por lo general necesitan un mínimo de 120 individuos de referencia para cada grupo (o subgrupo) que deba ser caracterizado.

Por ejemplo, para establecer un intervalo de referencia para hemoglobina — un examen que es dependiente de género — el laboratorio necesitaría obtener resultados de hemoglobina al menos de 240 individuos de referencia (120 hombres y 120 mujeres). Estos individuos son reclutados por lo general de la población regional general (esencialmente potenciales usuarios del servicio del laboratorio) y luego seleccionados para su inclusión en el estudio utilizando criterios cuidadosamente definidos. La selección se realiza con frecuencia aplicando un cuestionario para evaluar la condición de salud del individuo. Algunas veces también se requiere de un examen físico como parte del criterio de aceptación del individuo para el estudio.

El establecimiento de los intervalos de referencia requiere de una planificación, control y una documentación cuidadosos de cada aspecto del estudio. Así, los intervalos de referencia resultantes están bien caracterizados en términos de la variación atribuible a factores pre analíticos y analíticos. Estos protocolos formales son particularmente útiles cuando un laboratorio necesita establecer

sus propios intervalos de referencia para un examen particular. Esta situación podría ocurrir si un laboratorio ha modificado un método previamente aprobado por la FDA o ha desarrollado un examen propio. Desafortunadamente, estos protocolos son costosos y pueden resultar inaccesibles para establecimientos más pequeños. Incluso los grandes laboratorios encuentran cada vez más difícil llevar a cabo estos estudios integrales considerando su relación costo beneficio. Por esto, los laboratorios son cada vez más dependientes de los fabricantes para establecer científicamente los intervalos de referencia sólidos que puedan ser verificados utilizando acercamientos más simples, de menor costo y que impliquen menos trabajo.

En éste capítulo, el enfoque está en la *transferencia* de los intervalos de referencia, los cuales requieren considerablemente menos esfuerzo y menos datos que los necesarios para *establecer* intervalos de referencia. El intervalo de referencia de mayor interés durante el proceso de validación de métodos es el que describe los valores del examen observados típicamente en una población “sana”. Éste intervalo se llama históricamente “rango normal” y se deriva del examen de especímenes de individuos que cumplen con los criterios de “buena” salud (por ejemplo “no tener problemas de salud conocidos, ser ambulatorios, no estar en un régimen regular de fármacos, tener un peso dentro de las normas recomendadas, etc.”). Los resultados del examen (valores de referencia) para éste grupo son analizados estadísticamente para determinar un intervalo de valores que incluyan un porcentaje determinado de todos los valores (intervalo de referencia) de la muestra de la población. Por tradición, éste intervalo incluye un 95% de los valores (usualmente el 95% central). Un par de valores del examen (llamados límites de referencia superior e inferior) representan los extremos del intervalo. Los resultados de pacientes que caigan fuera de los límites de referencia así establecidos, se señalan generalmente como resultados “anormales”.

Acercamientos de transferencia a considerar

La Guía Aprobada de la CLSI C28-A3 [7] describe diferentes alternativas a las que puede acceder un laboratorio para validar la “transferencia” de intervalos de referencia establecidos a un laboratorio individual:

1. Juicio divino.

La aceptabilidad de la transferencia podría ser valorada subjetivamente en base a la consistencia entre “la demografía y la geografía” de la población empleada para realizar el estudio y la demografía de la población a examinar en el laboratorio. El laboratorio simplemente revisa la información entregada y verifica subjetivamente que los intervalos de referencia son aplicables a la población del laboratorio que los está adoptando y a los métodos de ensayo del propio laboratorio. Para hacer esto, toda la información sobre el estudio

original debería ser requerida y estar disponible en el laboratorio. Esto incluye los datos demográficos del grupo de referencia, el proceso de selección, las condiciones pre analíticas del estudio tales como la preparación del individuo y la recolección del espécimen y las técnicas de manejo, el sistema analítico utilizado, y el método estadístico utilizado para establecer los intervalos. En ocasiones es útil pedir los valores de referencia originales y analizarlos nuevamente para verificar el análisis estadístico original. La mayoría de los casos de transferencia involucran la adopción de intervalos de otro laboratorio utilizando el mismo sistema analítico o intervalos establecidos por el fabricante del método.

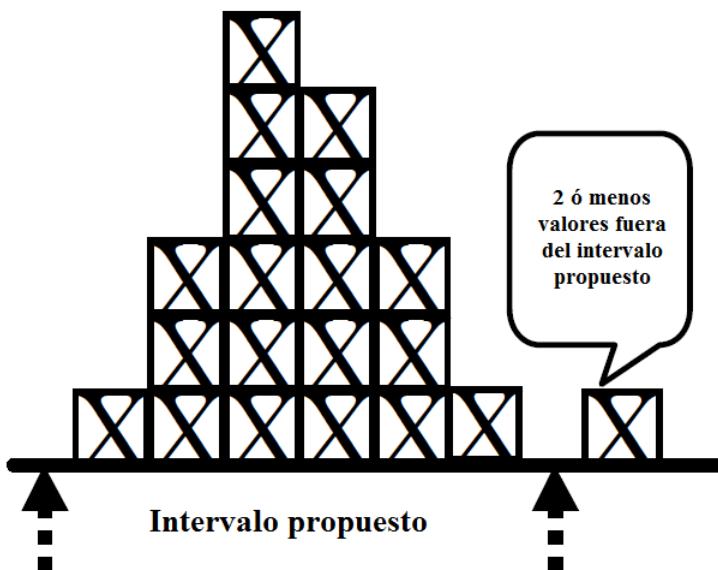
- Las regulaciones de la Regla Final de CLIA permiten que el Director Médico de un laboratorio haga ésta valoración y éste juicio.
- Mientras la transferencia requiere sólo la firma apropiada, el enfoque depende de una investigación cuidadosa de las recomendaciones publicadas, acceder a la información apropiada y experiencia significativa clínica y en el laboratorio para asegurarse de la comparabilidad de las condiciones.
- La provisión de los intervalos de referencia para sub-poblaciones, particularmente para intervalos pediátricos por intervalos de edad, a menudo requieren de éste acercamiento dada la dificultad para obtener especímenes suficientes para establecer experimentalmente o verificar este tipo de intervalos de referencia.

2. Verificación con 20 muestras.

La validación experimental podría ser realizada recolectando y analizando especímenes de 20 individuos que representan una muestra de la población de referencia. Si dos o menos resultados del examen caen fuera de los límites de referencia propuestos, **el intervalo de referencia se considera verificado**, como se ilustra en la figura.

- Experimentalmente, éste es un enfoque simple. Sólo requiere un mínimo de datos, y el estudio provee un criterio claro para interpretar los datos y verificar la transferencia.
- Es más fácil seleccionar la población adulta, hombres adultos o mujeres adultas como el intervalo de referencia a verificar que una sub-población más pequeña.

Verificación con 20 muestras



3. Estimación con 60 muestras.

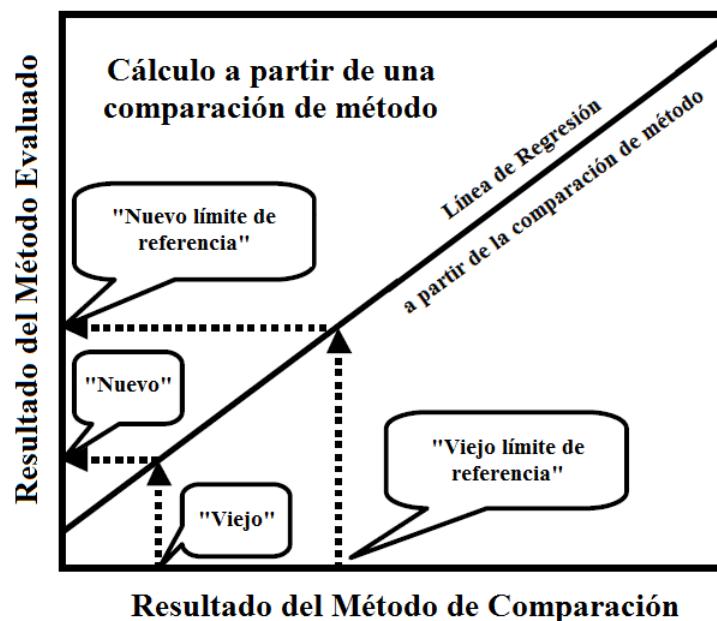
Una validación experimental podría ser realizada recolectando y analizando especímenes de 60 individuos que representan una muestra de la población de referencia. El intervalo de referencia se estima a partir de estos datos y se compara con el intervalo propuesto utilizando una fórmula estadística que compare las medias y las desviaciones estándar de las dos poblaciones [7].

- A veces encontrar incluso 60 individuos de referencia puede ser una tarea desalentadora. Cuando mucho, se deberían usar no menos de 40 individuos para calcular estadísticamente los intervalos de referencia. Cuando se utilizan menos de 40 individuos, es mejor reportar los descubrimientos en términos de los valores mínimos y máximos del ensayo observados a lo largo de un histograma mostrando la distribución de los valores.
- Las comparaciones estadísticas de los intervalos observados y los propuestos son más complicadas y por tanto menos atractivas para verificar un intervalo de referencia que el acercamiento de 20 muestras.

4. Cálculos a partir del método de comparación.

El documento de CLSI también reconoce — pero no respalda — otro acercamiento que ajustaría o corregiría los intervalos de referencia propuestos o reportados en base al sesgo metodológico observado y la relación matemática demostrada entre los métodos analíticos en uso (como se muestra en la figura

abajo). Las estadísticas de regresión obtenidas de un estudio de comparación de métodos podrían ser usadas para calcular los límites de referencia (X_{inf} y X_{sup}) para el nuevo método ($Y_{inf} = a + bX_{inf}$, $Y_{sup} = a + bX_{sup}$, donde a es la intersección-Y, y b es la pendiente de la línea de regresión).



- Éste enfoque es atractivo cuando el laboratorio ha realizado los estudios necesarios para establecer sus propios intervalos de referencia en el pasado y luego cambia a un nuevo método analítico. Los datos propios deberían, por lo tanto, estar disponibles para documentar la comparación del nuevo método con el antiguo.
- Se recomienda la verificación por análisis de 20 muestras si hay alguna duda sobre la confiabilidad de los intervalos de referencia a ser trasladados del método de comparación.
- La transferencia por cálculos debería estar limitada a un cambio de métodos para minimizar la “cascada” potencial de errores que se podrá generar al establecer los nuevos límites de referencia.
- El enfoque matemático también es apoyado por Koivula [8] para la transferencia de intervalos de referencia de un laboratorio de referencia a un laboratorio individual. El laboratorio analiza seis o siete muestras seleccionadas cuidadosamente de un programa de evaluación externa de la calidad y calcula la línea de regresión entre los valores del laboratorio individual y los valores del laboratorio de referencia. El intervalo de

referencia del laboratorio de referencia es usado entonces con la ecuación de regresión para calcular los límites de referencia del laboratorio individual.

¿Qué hacer?

Para exámenes donde hay intervalos de referencia bien establecidos para el método comparativo en su laboratorio, transfiera esos intervalos matemáticamente usando la ecuación de regresión obtenida del experimento de comparación de métodos realizado en su laboratorio. Asegúrese de que las estadísticas de regresión sean confiables siguiendo los lineamientos para realizar el experimento de comparación de métodos y generando el análisis estadístico de datos adecuado. La transferencia sólo requerirá unos pocos cálculos adicionales, haciendo rápido y fácil determinar los nuevos límites. Éste acercamiento debería ser apto para exámenes comunes de las áreas de química clínica y hematología donde se espera que haya una concordancia uno a uno.

Para exámenes donde hay diferencias sistemáticas entre el nuevo método y el de comparación, use el enfoque matemático, antes abordado, para estimar los intervalos de referencia y compararlos con el propuesto por el fabricante. Verifique posteriormente los límites así obtenidos por un análisis de 20 especímenes de individuos sanos. Éste enfoque sería apropiado con métodos enzimáticos donde existen a menudo diferencias proporcionales entre los métodos y los inmunoensayos donde cambios sistemáticos podrían ocurrir entre generaciones de sistemas de medición.

Use la estimación con 20 muestras para transferir los intervalos de referencia propuestos por los fabricantes cuando el fabricante provea la documentación de un estudio adecuado de valores de referencia.

Use la estimación con 60 muestras para hacer estimados del intervalo de referencia cuando la información del intervalo de referencia propuesto por el fabricante no sea adecuada, cuando el nuevo método esté basado en un principio de medición diferente con diferente especificidad de medida, o cuando el examen está siendo aplicado a una población de pacientes diferente.

Use el juicio divino cuando no haya datos experimentales que apoyen la transferencia de los intervalos de referencia.

Referencias

1. Solberg HE. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *Clin Chim Acta* 1987;167:111-118.
2. PetitClerc C, Solberg HE. Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem* 1987;25:639-644.
3. Solberg HE, PetitClerc C. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *Clin Chim Acta* 1988;177:S-S12.
4. Solberg HE, Stamm D. Approved recommendation on the theory of reference values. Part 4. Control of analytical variation in the production, transfer, and application of reference values. *Eur J Clin Chem Clin Biochem* 1991;29:531-535.
5. Solberg HE. Approved recommendations (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem* 1987;25:656-656.
6. Dybkaer R, Solberg HE. Approved recommendations (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. *J Clin Chem Clin Biochem* 1987;25:657-662.
7. CLSI C28-A3: How to define and determine reference intervals in the clinical laboratory - Second edition - Approved Guideline. Wayne, PA: Clinical Laboratory Standards Institute, 2000.
8. Koivula T. Possibilities for quality assurance of reference intervals. *Scand J Clin Lab Invest* 1995;55, Suppl. 222:17-20.

16: ¿Cómo juzgar el desempeño de un método?

Usted ya ha realizado los experimentos, tabulado los resultados, graficado los datos y calculado las estadísticas. Éste capítulo describe el siguiente paso, que es hacer una evaluación sobre desempeño del método. Se recomienda una técnica gráfica — La Carta de Decisión de Método — y se ilustran su construcción y aplicación.

Objetivos:

- Construir una Carta de Decisión de Métodos.
- Evaluar el desempeño de un método usando la Carta de Decisión de Métodos.
- Recomendar acciones apropiadas en base al desempeño del método observado.

Materiales del capítulo:

- **VM — La decisión sobre el desempeño de un método,**
por James O. Westgard, PhD
- **Caja de herramientas de datos para validación de métodos**
<http://www.westgard.com/mvtools.html>
- **Calculadora de puntos operativos normalizados**
<http://www.westgard.com/normcalc.htm>
- **Problema: Datos de validación de un método de colesterol.**

Cosas qué hacer:

- Estudiar los materiales.

Preguntas de Autoevaluación:

- ¿Cómo se incorpora un requisito de la calidad en una Carta de Decisión del Método?
- ¿Cuál es una fuente común de requisitos de la calidad para un ensayo?
- ¿Cuáles son los parámetros de desempeño críticos que se grafican en una Carta de Decisión del Método?
- ¿Cuáles son las clasificaciones de desempeño en una Carta de Decisión del Método?
- ¿Cuál es su juicio sobre un método de calcio cuya S observado es 0.2 mg/dL y cuyo Sesgo observado es 0.1 mg/dL a un nivel de decisión médica de 11.0?

Validación de Métodos: La Decisión sobre el Desempeño de un Método

James O. Westgard, PhD

Usted ya ha realizado los experimentos, tabulado los resultados, graficado los datos y calculado las estadísticas. Ahora tiene que decidir si el método es aceptable. ¿Cómo decide si el método es suficientemente bueno para ser usado en su laboratorio?

¿Cuál es el enfoque adecuado?

Recuerde el secreto mas interno oculto y profundo sobre la validación de métodos—se trata de EVALUACIÓN DEL ERRORES. La decisión sobre la aceptabilidad del desempeño de un método depende del tamaño de los errores observados en relación a algunos estándares o requisitos de la calidad que definen el error clínico permitido. El desempeño del método es aceptable cuando los errores observados son más pequeños que el error clínico permitido. El desempeño del método NO es aceptable cuando los errores observados son más grandes que el error clínico permitido.

Debería de hecho definir los errores clínicos permitidos a niveles de decisión médica relevantes al inicio para ayudarse a guiar el diseño de los experimentos y la recolección de los datos. Lo que quedará por hacer es comparar sus errores observados con los errores permitidos definidos desde un punto de vista clínico.

¿Cómo deberían asentarse los requisitos para errores clínicamente permitidos?

En la literatura científica, los requisitos para la calidad analítica han sido definidos en tres formatos diferentes—el error total permitido, el desvío estándar permitido y el sesgo permitido. Un error total permitido sienta los límites del efecto combinado de los errores aleatorio y sistemático de un método, donde un desvío estándar permitido y un sesgo permitido sientan los límites separados para los errores aleatorio y sistemático, respectivamente.

Parecerían ser útiles requisitos separados para el desvío estándar permitido y sesgo permitido dado que éstas estadísticas pueden ser calculadas directamente de los datos experimentales (por ej., un desvío estándar se calcula a partir de los datos de un experimento de replicación y un sesgo para partir de los datos los datos de un experimento de comparación de métodos). Sin embargo, la calidad del

resultado de un paciente se determina por el error neto o total, efecto combinado de ambos errores, por lo que el error total es clínicamente más relevante [1]:

“Los médicos piensan más bien en términos más relevantes desde un punto de vista clínico error analítico total, que incluyen tanto el componente aleatorio como el sistemático. Desde este punto de vista, todos los tipos de error analítico son aceptables en tanto que el error analítico total sea menor que una cantidad especificada. El error total es clínicamente más útil; después de todo, hace poca diferencia para un paciente si un valor de laboratorio tiene error por errores analíticos aleatorios o por sistemáticos, y finalmente él es el único que deberá vivir con ese error”.

¿Dónde encuentra recomendaciones para errores totales permitidos?

Una fuente común son las encuestas de programas de evaluación externa de la calidad o un esquema de evaluación de la competencia en los que participe su laboratorio. Estos programas definen generalmente un “valor blanco” central y un rango de valores alrededor de ese blanco que son considerados como aceptables. Dado que estos programas usualmente solicitan procesar el espécimen una única vez en cada encuesta, ese único resultado estará afectado por el error aleatorio y el error sistemático de su procedimiento de medida. El “rango aceptable” es por tanto un requisito de desempeño analítico en el formato de un error total permitido.

Para los laboratorios de E.U.A., CLIA tiene disponible un listado de error total permitido basado en los criterio de error total aceptable para esquemas de evaluación de la competencia, que ha sido publicado en el Registro Federal y provee recomendaciones para más de 70 exámenes diferentes. Vea la lista de criterios provistos en el Apéndice 1. Estos criterios son presentados de 3 formas diferentes:

- Como un límite de concentración absoluta, por ej., valor blanco \pm 1 mg/dL para calcio;
- Como un porcentaje, por ej., valor blanco \pm 10% para albúmina, colesterol y proteínas totales;
- Como el rango determinado de un grupo de encuestas, por ej., el valor blanco \pm 3 desviaciones estándar para la hormona estimulante de la tiroides (TSH, por sus siglas en Inglés).

En algunos pocos casos, se dan dos conjuntos de límites, por ej., el requisito de la calidad para la glucosa es el valor blanco \pm 6 mg/dL ó \pm 10%, cualquiera que sea mayor. A un nivel de decisión médica de 50 mg/dL, el error total permitido

es 6 mg/dL ó 12%. A un nivel de decisión médica de 125 mg/dL, el error total permitido es 10% ó 12.5 mg/dL.

¿Cómo se comparan los errores observados con el error total permitido?

Para estimar el error aleatorio del método a partir del experimento de replicación usted calculará una desviación estándar (S) o un coeficiente de variación (CV). Para estimar el error sistemático a partir del experimento de comparación de métodos usted calculará el sesgo entre las medias obtenidas por los métodos de examen y comparativo, o usará estadísticas de regresión para calcular la diferencia esperada a niveles particulares de decisión médica. Estas estimaciones de errores aleatorio y sistemático necesitan ser combinadas para juzgar su efecto total.

La literatura provee cuatro recomendaciones diferentes sobre cómo combinar los errores aleatorio y sistemático:

- Añada Sesgo + 2 veces la S observada [ref 1], esto es, Sesgo + 2S < TE_a;
- Añada Sesgo + 3 veces la S observada [ref 1], esto es, Sesgo + 3S < TE_a;
- Añada Sesgo + 4 veces la S observada [ref 1], esto es, Sesgo + 4S < TE_a;
- Añada Sesgo + 5 veces la S observada [ref 1], esto es, Sesgo + 5S < TE_a.

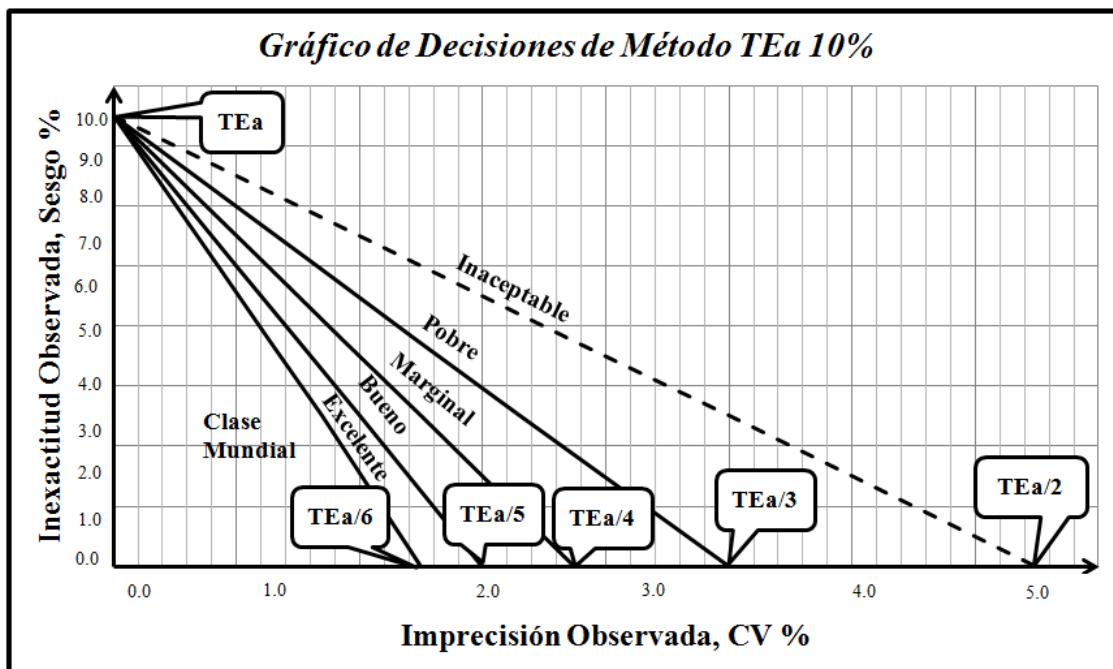
Más que escoger entre estas recomendaciones, todas pueden ser incorporadas en una herramienta de decisión gráfica — una Carta de Decisión del Método. La carta es fácil de construir, minimiza la necesidad de cálculos adicionales y provee una imagen gráfica que simplifica la interpretación y evaluación del desempeño del método.

¿Cómo se construye una Carta de Decisión del Método?

Primero, exprese el error total permitido como un porcentaje de la concentración de decisión médica evaluada. La mayoría de los errores permitidos de CLIA ya se expresan de ésta forma. Para aquellos errores permitidos que están expresados en unidades de concentración, exprese el error permitido como un porcentaje de la concentración de decisión médica de interés, esto es, divida el error permitido por la concentración de decisión médica y multiplíquelo por 100 para expresarlo como porcentaje.

Luego, tome una hoja o papel para graficar y haga lo siguiente:

1. Rotule el eje de las ordenadas (Y) como “Inexactitud permitida, (Sesgo, %)” y establezca una escala desde 0 hasta el TE_a , por ej., para un $TE_a = 10\%$, establezca una escala en el eje de las ordenadas de 0 a 10% en incrementos de 1%.
2. Rotule el eje de las abscisas (X) como “Imprecisión permitida, (CV, %)” y establezca una escala desde 0 hasta $0.5 TE_a$, por ej., para un $TE_a = 10\%$, establezca una escala en el eje de las abscisas de 0 a 5% en incrementos de 0.5%.
3. Dibuje una línea para el Sesgo + 2S desde el valor del TE_a en el eje de las ordenadas (Y) hasta el valor de $0.5 TE_a$ en el eje de las abscisas (X), por ej., para un $TE_a = 10\%$, dibuje la línea del 10% (eje y) al 5% (eje x).
4. Dibuje una línea para el Sesgo + 3S desde el valor del TE_a en el eje de las ordenadas (Y) hasta el valor de $0.33 TE_a$ en el eje de las abscisas (X), por ej., para un $TE_a = 10\%$, dibuje la línea del 10% (eje y) al 3.33% (eje x).
5. Dibuje una línea para el Sesgo + 4S desde el valor del TE_a en el eje de las ordenadas (Y) hasta el valor de $0.25 TE_a$ en el eje de las abscisas (X), por ej., para un $TE_a = 10\%$, dibuje la línea del 10% (eje y) al 2.5% (eje x).
6. Dibuje una línea para el Sesgo + 5S desde el valor del TE_a en el eje de las ordenadas (Y) hasta el valor de $0.20 TE_a$ en el eje de las abscisas (X), por ej., para un $TE_a = 10\%$, dibuje la línea del 10% (eje y) al 2.0% (eje x).
7. Dibuje una línea para el Sesgo + 6S desde el valor del TE_a en el eje de las ordenadas (Y) hasta el valor de $0.17 TE_a$ en el eje de las abscisas (X), por ej., para un $TE_a = 10\%$, dibuje la línea del 10% (eje y) al 1.7% (eje x).

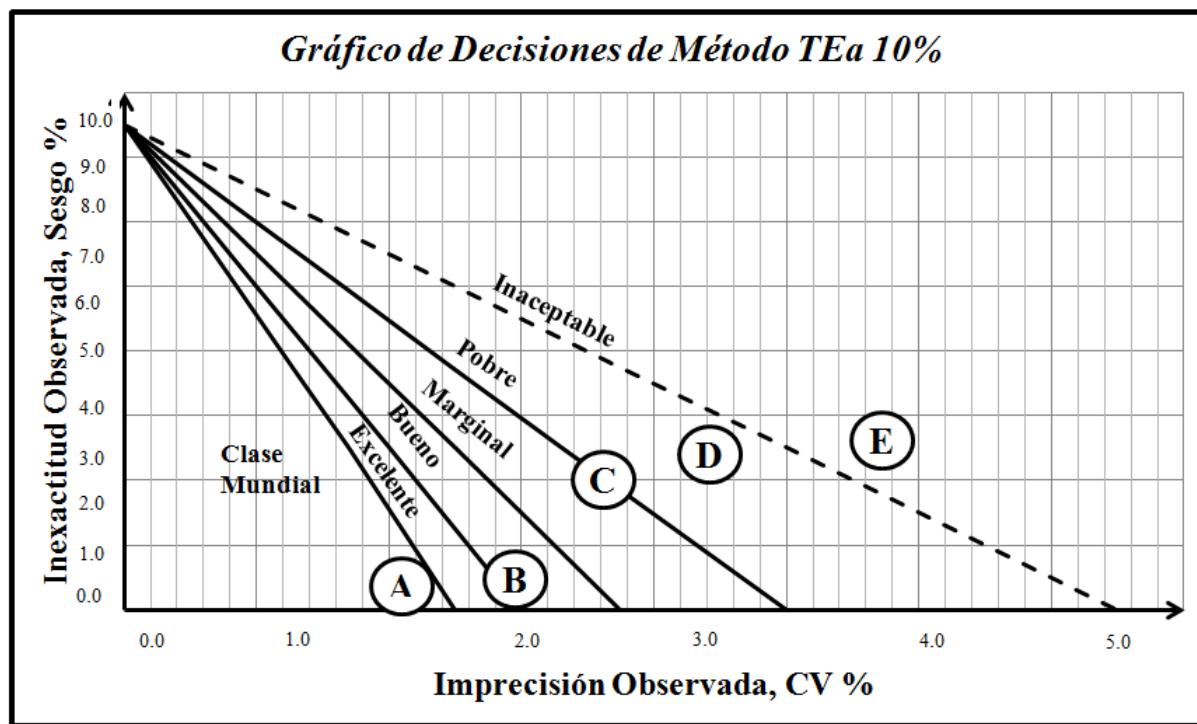


8. Marque las regiones definidas entre las líneas trazadas como: “inaceptable”, “pobre”, “marginal”, “bueno”, “excelente” y “clase mundial” como se muestra en la figura.

¿Cómo se usa la Carta de Decisión del Método?

Expresé su desvío estándar y sesgo observados para el procedimiento de medida evaluado en porcentaje, luego grafique el punto cuya coordenada en “x” sea la imprecisión observada y su coordenada en “y” la inexactitud observada. A éste punto se le llama el “punto operativo” pues describe cómo opera su método. Evaluará el desempeño de su método en base a la locación del punto operativo, como sigue:

- Un método con **desempeño inaceptable** no cumple con sus requisitos de la calidad, incluso cuando el método está trabajando adecuadamente. No es aceptable para operaciones de rutina.
- Un método con **desempeño pobre** podría haber sido considerado adecuado antes de la introducción reciente de los principios de Gestión de la Calidad Seis Sigma, pero los estándares industriales de referencia sientan ahora un estándar mínimo de desempeño 3-Sigma para un proceso de producción de rutina, por lo que el desempeño en la región entre 2-Sigma y 3-Sigma no es hoy en día verdaderamente satisfactorio.
- Un método con **desempeño marginal** provee la calidad necesaria cuando todo está trabajando correctamente. Sin embargo, será difícil de manejar en operaciones de rutina, requerirá de 4 a 8 controles por corrida y una estrategia de Control de la Calidad Total que enfatiza en operadores bien entrenados, rotación reducida del personal, mantenimiento preventivo más agresivo, monitoreo cuidadoso de los resultados de los pacientes y esfuerzos continuos para mejorar el desempeño del método.
- Un método con **desempeño bueno** cumple con sus requisitos de la calidad y puede ser bien manejado en operaciones de rutina con mediciones de 2 a 4 controles por corrida usando procedimientos de Control de la Calidad de reglas múltiples o una única regla de control con límites de control a 2.5S.
- Un método con **desempeño excelente** es claramente aceptable y puede ser bien manejado en operaciones de rutina con sólo 2 controles por corrida analítica con una sola regla de control con límites de control de 2.5S ó 3.0S.
- Un método con **desempeño de clase mundial** es incluso más fácil de manejar y controlar, usualmente requieren sólo 1 ó 2 controles por corrida analítica y una sola regla de control con límites amplios, tales como 3.0S ó 3.5S.



Ejemplos de aplicaciones

Todos estos ejemplos ilustran la evaluación para métodos de colesterol, donde el requisito de la calidad de CLIA para desempeño aceptable es un error total permitido de 10%.

- Un método de colesterol con un CV de 1.5% y un Sesgo de 0.0% provee una calidad de clase mundial, como se muestra por el punto operativo marcado como “A”, cuya coordenada en “x” es 1.5 y en “y” es 0.0. Éste método es claramente aceptable y será fácil de manejar y controlar en operaciones de rutina usando 2 controles por corrida analítica y recurriendo a una sola regla de control con límites a 3.5S.
- Un método de colesterol con un CV de 2.0% y un Sesgo de 0.0% provee un desempeño excelente, como se muestra por el punto operativo “B”. Éste método es claramente aceptable y será controlable en servicios de rutina usando 2 controles en una sola corrida analítica y recurriendo a una sola regla de control con límites de control a 3.0S o 2.5S.
- Un método de colesterol con un CV de 2.0% y un Sesgo de 2.0% tiene un punto operativo que cae en la zona enmarcada por las líneas de buen desempeño y desempeño marginal, como se muestra en el punto operativo “C”. Se requiere una valoración cuidadosa del control estadístico interno

de la calidad y mostrará que se necesita un procedimiento de reglas múltiples con un total de 4 controles para garantizar la calidad deseada para éste método. Vea también la aplicación Control de la Calidad (QC, por sus siglas en Inglés) en Westgard Web en <http://www.westgard.com/qcapp3.htm>.

- D. Un método de colesterol con un CV de 3.0% y un Sesgo de 3.0% satisface las especificaciones del *National Cholesterol Education Program* (NCEP, por sus siglas en Inglés) [7]. Para valorar si éstas especificaciones de desempeño son adecuadas, se puede graficar un punto operativo con una coordenada en “y” de 3.0% y una coordenada en “x” de 3.0%, como se muestra en el punto operativo marcado como “D” en la figura acompañante. Tal método tiene un desempeño pobre y sólo alcanzará la calidad deseada *si todo está trabajando perfectamente*. Será muy difícil detectar problemas y mantener la calidad deseada durante el servicio de rutina [Note que discordamos con el CDC (*Centers for Disease Control and Prevention*) en la metodología apropiada para asentar especificaciones de desempeño. La referencia [8] provee algo de discusión y debate. La referencia [9] es una defensa extensa de la metodología de CDC-NCEP].
- E. Un método de colesterol con un CV de 4.0% y un Sesgo de 3.0% podría ser representativo del tipo de métodos de tamizaje encontrados en supermercados y farmacias. Como se muestra en el punto operativo “E”, tales métodos no proveen la calidad necesaria para cumplir los requisitos de la calidad de CLIA para desempeño aceptable *incluso si todo está trabajando perfectamente*.

¡Inténtelo!

Defina el error total permitido para su examen. Construya una Carta de Decisión de Método usando una página o papel para graficar. Grafique la inexactitud observada (% sesgo) del experimento de comparación de métodos contra la imprecisión observada (% S o CV) del experimento de replicación. Vea donde está localizado éste punto operativo y juzgue si quiere o no implementar el método para servicios de rutina.

¡Pruebe con las Calculadoras en Línea!

En *Westgard Web*, están disponibles dos herramientas:

Method Decision Calculator

<http://www.westgard.com/mvtools.html>

Normalized Operating Point Calculator

<http://www.westgard.com/normcalc.htm>

La Calculadora *Method Decision Calculator* trabaja enviándole los valores que usted introduce al servidor para el cálculo y la preparación de la Carta de Decisión del Método. La Calculadora *Normalized Operating Point Calculator* es una calculadora *Javascript* que realiza los cálculos en su computadora. Sin embargo, no puede realizar un gráfico así que necesita imprimir la Carta de Decisión del Método Normalizada y graficar manualmente su punto operativo.

También está disponible una Carta de Decisión del Método automatizada en *Westgard Web* — como una hoja de cálculo de Excel. Puede descargar el archivo en <http://www.westgard.com/downloads/>

Referencias:

1. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974;20:825-33.
2. U.S. Department of Health and Social Services. Medicare, Medicaid, and CLIA Programs: Regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final Rule. *Fed Regist* 1992(Feb 28);57:7002-7186.
3. Ehrmeyer SS, Laessig RH, Leinweber JE, Oryall JE. 1990 Medicare/ CLIA final rules for proficiency testing: Minimum interlaboratory performance characteristics (CV and Bias) needed to pass. *Clin Chem* 1990;36:1736-40.
4. Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. *Clin Chem* 1990;36:1629-32.
5. Westgard JO. Six Sigma Quality Design and Control, 2nd Edition, Chapter 7. Madison WI: Westgard QC, Inc., 2006, pp.91-105.
6. Westgard JO. A method evaluation decision chart (MEDx Chart) for judging method performance. *Clin Lab Science*. 1995;8:277-83.
7. Westgard JO, Wiebe DA. Cholesterol operational process specifications for assuring the quality required by CLIA proficiency testing. *Clin Chem* 1991;37:1938-44.
8. Westgard JO, Wiebe DA. Adequacy of NCEP recommendations for total cholesterol, triglycerides, HDLC, and LDLC measurements. *Clin Chem* 1998;44:1064-1066.
9. Caudill SP, Cooper GR, Smith SJ, Myers GL. Assessment of current National Cholesterol Education Program guidelines for total cholesterol, triglycerides, HDL-cholesterol, and LDL-cholesterol measurements. *Clin Chem* 1998;44:1650-8.

17: ¿Cuál es un procedimiento práctico para validar un método?

Este capítulo revisa el enfoque general de la validación de métodos y provee un conjunto de hojas de cálculo para recolectar las cantidades mínimas de datos necesarios en cada experimento. Si usted decide trabajar con este modelo y emplear estas hojas de cálculo, usted podrá planificar y organizar su propio estudio de validación de métodos.

Objetivos:

- Revisar el enfoque general de validación de métodos.
- Desarrollar hojas de cálculo para recolectar los datos experimentales.
- Organizar éstas hojas de cálculo en un plan práctico de validación de métodos.
- Determinar el apoyo estadístico necesario para implementar un plan de validación de métodos.

Materiales del capítulo:

- VM – Las aplicaciones del mundo real,
por James O. Westgard, PhD
- VM – Las hojas de cálculo

Cosas qué hacer:

- Estudiar los materiales.
- Adaptar las hojas de cálculo para ajustarlas a sus necesidades.

Preguntas de Autoevaluación:

- ¿Cuáles son los experimentos mínimos necesarios para sus exámenes?
- ¿Cuál es el número mínimo de mediciones necesarias para cada experimento?
- ¿Cuál es la mejor forma de organizar estos experimentos?
- ¿Cuál es la forma más práctica de hacer los cálculos a partir de los datos?

Validación de Métodos: Las Aplicaciones del Mundo Real

James O. Westgard, PhD

Los capítulos previos le han otorgado habilidades para establecer su propio proceso de validación de métodos. Es importante establecer un proceso sistemático a la medida de las necesidades de su laboratorio y de las características del método que va a ser examinado. ¡No asuma que el desempeño del método está bien sólo porque ha comprado un nuevo instrumento o un nuevo lote de reactivos! En el mundo real, los métodos todavía tienen problemas. ¿Recuerda los *Mitos de la calidad*?

Una revisión del proceso de Validación de Métodos

Aquí hay un breve resumen de nuestras lecciones sobre validación de métodos. Puede revisar los capítulos anteriores para obtener mayores detalles y aclarar los puntos más importantes.

- La Validación de Métodos debería ser un proceso estándar en el laboratorio clínico, pero el proceso no necesita ser exactamente el mismo para cada laboratorio o para cada método validado por un laboratorio. Vea *Gestión de la Calidad – La necesidad de procesos estandarizados y estándares para la calidad* para una visión general del proceso de manejo de la calidad que se necesita en los laboratorios clínicos y el rol de la validación de métodos al establecer los procesos estandarizados de examen.
- Recuerde que el propósito de la validación de métodos es la valoración del error. Vea *El significado secreto mas interno, oculto y profundo* para una descripción de los errores aleatorios, sistemático y total que son el foco de los estudios de validación de métodos.
- Note que a un laboratorio de E.U.A. se le requiere por las regulaciones de CLIA que “demuestre que antes de informar resultados de pacientes, puede alcanzar las especificaciones de desempeño para exactitud, precisión y rango reportable, comparables a aquellas establecidas por el fabricante. El laboratorio también debe verificar que el rango de referencia del fabricante es adecuado para la población de pacientes del laboratorio”. Para métodos *non-waived*, CLIA también requiere la verificación de la sensibilidad analítica y de la especificidad analítica. Vea *Las Regulaciones*.
- Otros factores o características importantes para el método, tales como

costo del examen, tipos de espécimen, volúmenes del espécimen, tiempo requerido para el análisis, proporción del análisis, equipo requerido, personal requerido, eficiencia, seguridad, etc., deben ser considerados durante la selección del método analítico. Vea *Seleccionando un método para validar*.

- El enfoque en la validación de métodos es realizar una serie de experimentos diseñados para estimar ciertos tipos de errores analíticos, por ej., un experimento de linealidad para determinar rango reportable, un experimento de replicación para estimar imprecisión o error aleatorio, un experimento de comparación de métodos para estimar inexactitud o error sistemático, o experimentos de interferencia y recuperación para estimar específicamente errores sistemáticos constantes y sistemáticos proporcionales (especificidad analítica), y un experimento de límite de detección para caracterizar la sensibilidad analítica. Vea *El Plan Experimental*. Para detalles sobre los diferentes tipos de experimentos, vea lo siguiente:
 - *El experimento de linealidad o rango reportable.* Deberían analizarse un mínimo de 5 especímenes con valores conocidos o asignados por triplicado para evaluar el rango reportable.
 - *El experimento de replicación.* Se recomienda hacer la determinación de 20 replicados de al menos dos niveles de materiales de control para estimar la imprecisión o el error aleatorio del método.
 - *El experimento de comparación de métodos.* Deberían analizarse un mínimo de 40 especímenes de pacientes por el nuevo método (método analítico) y un método establecido (método de comparación) para estimar la inexactitud o el error sistemático del método.
 - *Los experimentos de interferencia y recuperación.* Se analizan usualmente las interferencias comunes, tales como lipemia, hemólisis e ictericia, junto con las interferencias específicas del ensayo y la metodología. Los experimentos de recuperación se utilizan para medir las interferencias competitivas, como pueden ser los posibles efectos de proteínas y metabolitos en los especímenes.
 - *El experimento de límite de detección.* El límite de detección está indicado para aquellos métodos cuyos valores analíticos bajos tienen una interpretación fundamental desde el punto de vista clínico. Generalmente, se analizan 20 veces un espécimen “blanco” y un espécimen “suplementado” con la cantidad de analito declarada por el fabricante en el inserto para el límite de detección.

- Los datos recolectados en los diferentes experimentos necesitan ser resumidos (por cálculos estadísticos) para brindar estimaciones de los errores analíticos que son el foco de cada experimento. Vea *Conjunto de herramientas para el análisis de datos*.
- La aceptabilidad de estos errores observados se juzga por comparación con los estándares de la calidad, esto es, recomendaciones para el error total permitido tales como las provistas por los criterios de los esquemas de evaluación de la competencia de CLIA. El desempeño del método es aceptable cuando los errores observados son menores que los límites establecidos para error permitido del procedimiento de medida evaluado. Una herramienta gráfica – la Carta de Decisión de Métodos – puede ser utilizada para clasificar el desempeño como clase mundial (Seis Sigma), excelente, bueno, marginal, pobre o inaceptable. [1] Vea *La decisión del desempeño del método*.
- Si se juzga aceptable el desempeño de un método, se deben verificar los intervalos de referencia. Vea *Transferencia de Intervalos de Referencias*.

Aplicaciones con datos publicados

Una revisión crítica de la literatura siempre es un buen punto de partida al seleccionar y evaluar un método. Esto puede incluir artículos científicos así como las descripciones del método del fabricante. El tiempo y esfuerzo necesarios para los estudios de validación del método en su laboratorio pueden ser minimizados con una valoración cuidadosa de los datos en la literatura.

Los estudios publicados de validación de métodos rara vez siguen un proceso de validación estandarizado. Por lo tanto, es necesario imponer su propio sistema de organización, de análisis y de interpretación de datos si quiere que sus resultados tengan sentido. Este es un proceso de revisión crítica, que no implica sólo aceptar el formato, los datos y el análisis y las conclusiones que han sido publicadas.

- Defina el requisito de la calidad en forma de un error total permitido (TE_a) para el examen (o exámenes) de interés a concentraciones próximas o que representen niveles de decisión médica críticos para la interpretación de los exámenes. Note que solo unas pocas revistas científicas requieren a los autores declarar la calidad que ellos consideran aceptable, por lo que las conclusiones de muchos estudios publicados rara vez hacen referencia a algún estándar o requisito de la calidad. La excepción notable es la revista *Clinical Chemistry* que comenzó en enero de 1999 a aconsejar a los escritores contribuyentes a dar referencia de sus descubrimientos de desempeño del método para estándares de la calidad definidos [2]. Aún así, pocos artículos siguen éste consejo.

- Prepare una “página de datos” para resumir la información sobre los experimentos. Liste los experimentos estándar que esperaría encontrar, por ej., el rango reportable, la replicación intradía (repetibilidad), la replicación interdía (Intermedia o intralaboratorio), la interferencia, la recuperación y la comparación de métodos.
- Vea el informe publicado para localizar los diferentes experimentos, resuma los factores fundamentales (número de especímenes de pacientes, número de replicados de mediciones, etc.), e identifique las fortalezas y debilidades de los estudios publicados.
- Para un *experimento de replicación*, valore la adecuación de las concentraciones de los materiales de control, la matriz de la muestra, el periodo de tiempo del estudio y otras condiciones, tales como la cantidad de lotes de reactivo distintos incluidos, el número de analistas involucrados, etc. Tabule el número de mediciones, la media y la desviación estándar o el coeficiente de variación de cada material.
- Para un *experimento de interferencia*, valore si las sustancias y las concentraciones analizadas son apropiadas. Tabule la diferencia promedio o el sesgo al estimar el error sistemático constante.
- Para un *experimento de recuperación*, determine qué cálculos se realizaron (si se calculó la recuperación de la cantidad total medida o de la cantidad añadida, siendo la última la forma correcta). Tabule el número de experimentos y la recuperación promedio. Calcule el error proporcional (100–promedio %Recuperación), luego multiplique por las concentraciones críticas de decisión médica para estimar el error sistemático proporcional.
- Para un *experimento de límite de detección*, identifique claramente la definición del término particular utilizado y el enfoque experimental para hacer la estimación. Identifique las muestras analizadas, el número de replicados de medición y la ecuación para calcular el límite de detección.
- Para un *experimento de comparación*, valore si el método de comparación por sí mismo es una buena elección. Tabule el número de especímenes de pacientes analizados por los dos métodos, el rango de concentración estudiado y la distribución de los datos a lo largo de ese rango. Tabule los resultados estadísticos (con estadísticos “t” y estadísticas de regresión). Valores si las estadísticas de regresión lineal brindan estimaciones de error confiables inspeccionando el gráfico de comparación y también del valor del coeficiente de correlación (que debería ser 0.99 o mayor). Cuando las estadísticas de regresión son confiables, estime la inexactitud o el error sistemático a partir de la ecuación $ES = (a + bX_c) - X_c$, donde “a” es la intersección-Y, “b” es la pendiente, y “ X_c “ es la concentración

crítica de decisión médica. Si las estadísticas de regresión ordinarias no proveen las estimaciones confiables de los errores, determine si el sesgo de los estadísticos “t” sería confiable, lo que requiere que la media de los resultados de comparación sean cercanos a la concentración de decisión médica de interés.

- Use la Carta de Decisión del Método para evaluar si el desempeño del método es satisfactorio para su laboratorio. Muestre las estimaciones individuales de error sistemático o inexactitud (de los experimentos de interferencia y recuperación) como puntos en el eje de las ordenadas (Y); muestre las estimaciones individuales de error aleatorio o imprecisión como puntos en el eje de las abscisas (X). Valore el efecto combinado de los errores aleatorio y sistemático graficando los puntos operativos cuya coordenada en “y” es el sesgo o ES del experimento de comparación de métodos y cuya coordenada en “x” es el CV del estudio de replicación interdía.
- Revise las conclusiones y las recomendaciones de los autores. Resuelva las diferencias entre sus conclusiones y las de los autores. Identifique los factores que serán fundamentales si analiza el método en su propio laboratorio.

Aplicaciones en su propio laboratorio

Es importante tener un entendimiento claro del proceso de validación de métodos y estar bien organizado al llevar a cabo los estudios experimentales. Un buen mantenimiento de los registros es esencial para documentar las condiciones de los estudios (números de lote de reactivos, números de lote de los calibradores, recalibraciones, procedimientos de mantenimiento preventivo, cualquier cambio en el método o acción correctiva).

- Especifique cuidadosamente la aplicación, metodología y requisitos de desempeño para el ensayo de interés. Establezca el requisito de la calidad para el examen en forma de error total permitido (TE_a), como se especifica en muchos esquemas para la evaluación d la competencia. Conduzca una búsqueda cuidadosa en la literatura y seleccione el método que tiene la aplicación y las características metodológicas adecuadas y tiene una buena probabilidad de alcanzar el desempeño deseado.
- Desarrolle un plan de validación en base a las características del examen y del método que serán fundamentales para una aplicación exitosa en su laboratorio. Identifique los experimentos, especifique la cantidad de datos a ser recolectados e identifique las concentraciones de decisión médica o los rangos analíticos donde deben ser recolectados los datos. Establezca un horario para que el personal pueda llevar a cabo los estudios de validación.

- Implemente el plan de validación de métodos preparando un conjunto de hojas de cálculo que definan la cantidad de datos a ser recolectados en los diferentes experimentos. Estas hojas formalizarán la planificación de los experimentos y también facilitarán la recolección de los datos.
 - La *hoja de cálculo de rango reportable* debería tener columnas para los datos, el analista, la identificación de la muestra, el valor conocido o asignado, el resultado observado #1, el #2, el #3, el resultado promedio y los comentarios. El número de filas dependerá del número de especímenes analizados, que estará por lo general entre 5 y 10. También incluya información sobre la fuente de los especímenes, la preparación de los *pools* de especímenes, el fabricante y los números de lote de los materiales comerciales. Vea página 217.
 - La hoja de *cálculo de replicación* debería tener columnas para los datos, el analista, el resultado del material 1, del material 2, (del material 3 si es necesario), y los comentarios. El número de filas debería ser por lo menos de 20. También incluya información acerca del fabricante y los números de lote de los materiales de control analizados. Note que usualmente necesitará una hoja de cálculo para el estudio de replicación intradía (repetibilidad) y un segundo estudio interdía (intermedio o intralaboratorio). Vea página 218.
 - La *hoja de cálculo de comparación de métodos* debería tener columnas para la fecha, el analista, el número de identificación del espécimen, el resultado del examen (valor y), el resultado del examen por el método de comparación (valor x), la diferencia (y-x), y los comentarios. Añada columnas extra si se realizarán duplicados de las mediciones. El número de filas debería ser por lo menos de 40. Vea páginas 219-220.
- Comience a graficar los datos del estudio de comparación diariamente conforme se van obteniendo. Identifique valores discrepantes y repita esos exámenes por ambos métodos. Un gráfico de diferencia remarcará éstas discrepancias más claramente que un gráfico de comparación, pero se puede usar cualquiera o ambos para éste propósito.
- Realice los cálculos estadísticos que son apropiados para los datos recolectados en los diferentes experimentos.
- Use la Carta de Decisión de Método para valorar si el desempeño del método es satisfactorio para su laboratorio.
- Documente los estudios de validación de métodos. Si el desempeño del método es aceptable, prepare un procedimiento para documentar el proceso de validación estandarizado. Prepare materiales de docencia

para capacitaciones. Seleccione los materiales de control, las reglas de control y el número de mediciones de los controles adecuadamente para monitorear el desempeño en condiciones de rutina.

Adaptaciones para aplicaciones de laboratorios individuales

Debemos reconocer que la situación de cada laboratorio podría ser diferente, por lo tanto, se pueden realizar diferentes adaptaciones en diferentes laboratorios. El enfoque que proponemos es mantener los principios del proceso de validación de métodos, mientras hacemos el trabajo experimental tan eficiente y tan práctico como sea posible. Se presentan algunas ideas aquí en relación al alcance de los estudios, las habilidades personales y las técnicas de análisis de datos.

- El alcance de los estudios podría ajustarse a la información disponible en la literatura científica. Se puede realizar un trabajo mínimo cuando se han publicado estudios minuciosos. Siempre realice un experimento de linealidad o rango reportable y un estudio de replicación por lo menos durante diez días. Reduzca el número de pacientes a emplear en el estudio de comparación de métodos a 20 especímenes teniendo el cuidado de cubrir el rango reportable de manera homogénea con las diferentes concentraciones de los especímenes seleccionados. Minimice el uso de los estudios de recuperación y de interferencia. Del mismo modo, cuando reemplace un método o un instrumento con el mismo método o instrumento, o similar, su experiencia previa en el laboratorio le permitirá reducir la cantidad de datos necesarios para validar el nuevo método o instrumento.
- Nuevas tecnologías o cambios en los principios del método o de medición requerirán estudios de validación más extensivos. Se debe evaluar críticamente las nuevas metodologías que apenas están saliendo y no han sido muy difundidas. Si el laboratorio está involucrado en “exámenes de campo” de un fabricante, se requerirán estudios incluso más extensivos, mucho más allá de los mínimos sugeridos aquí para los estudios de validación básica de métodos.
- El personal del laboratorio involucrado en los estudios de validación de métodos puede tener una experiencia heterogénea (variable). Sin embargo, es importante tener al menos un analista con las habilidades para organizar los estudios, especificar la cantidad de datos a ser recolectados, monitorear los datos durante su recolección para identificar problemas obvios del método, inspeccionar cuidadosamente los datos para identificar los resultados discrepantes, analizar apropiadamente los datos estadísticos, interpretar críticamente los resultados y hacer cualquier cambio o ajuste

necesario al plan de validación. Otros analistas podrían llevar a cabo los experimentos y tabular los datos. La participación de varios analistas va a generar estimaciones más realistas de la imprecisión esperada bajo las condiciones de operación de rutina del método.

- El análisis de los datos debe ser comprendido por los analistas del laboratorio, si la información no es correctamente interpretada, las decisiones sobre el desempeño del método pueden ser incorrectas. Los datos que surgen de la comparación de métodos son más difíciles de analizar. Debe prepararse siempre un gráfico a partir de los datos. Se prefieren generalmente las estadísticas de regresión, pero los estadísticos “t” podrían ser suficientes cuando se obtiene la estimación del sesgo a una media que es muy cercana al nivel de decisión médica. Vea los capítulos 8, 12 y 18 para lineamientos más detallados del análisis de datos.

Referencias:

1. Westgard JO. A method evaluation decision chart (MEDx) Chart for judging method performance. Clin Lab Science 1995;8:277-83.
2. Information for authors. Journal of Clinical Chemistry
http://www.clinchem.org/info_ar/anal_meth.shtml

Referencias en Línea

A method evaluation decisión chart for judging method performance
<http://www.westgard.com/downloads/>

Method Decision Chart spreadsheet
<http://www.westgard.com/downloads/>

Points of Care in using statistics in method comparison studies
<http://www.westgard.com/essay19.htm>

Experimento

Rango Reportable

Ensayo: _____

Método: _____

Materiales: _____

#	Fecha	Analista	ID Muestra	Valor Asignado	Resultado 1	Resultado 2	Resultado 3	Promedio	Comentarios
1									
2									
3									
4									
5									
6									

**Experimento
de Replicación**

Ensayo: _____

Método: _____

Materiales: _____

#	Fecha	Hora	Analista	Resultado Material 1	Resultado Material 2	Resultado Material 3	Comentarios
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

**Experimento
Comparación Parte I**

Ensayo: _____

Método: _____

Materiales: _____

#	Fecha	Analista	ID Muestra	Resultado de Método Evaluado (y)	Resultado de Método de Comparación (x)	Dif (y-x)	Comentarios
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

Experimento

Comparación Parte II

Ensayo: _____

Método: _____

Materiales: _____

#	Fecha	Analista	ID Muestra	Resultado de Método Evaluado	Resultado de Método de Comparación (x)	Dif (y-x)	Comentarios
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							

18: ¿Cómo usa la estadística en el Mundo Real?

Este capítulo le otorgará una última dosis de estadística. La mayoría de los estudiantes aprenden estadística lentamente, unas pocas ideas a la vez. Toma varias dosis incrementar los conocimientos de estadística de una persona y mejorar sus habilidades para interpretar la estadística correctamente. Algunas veces se trata de escuchar o ver la información presentada de diferente forma. Esta es otra forma de trabajar a través del uso correcto y una interpretación apropiada de las estadísticas generadas a partir de los estudios de comparación de métodos.

Objetivos:

- Repasar el uso de la estadística en los estudios de comparación de métodos.
- Resumir consideraciones importantes para el uso apropiado de las estadísticas de regresión.
- Identificar cuando es necesario utilizar técnicas de regresión más sofisticadas como la regresión de Deming o el modelo de Passing-Bablok.

Materiales del capítulo:

- **Aspectos a considerar en el uso de las estadísticas en los estudios de comparación de métodos**, por James O. Westgard, PhD

Cosas qué hacer:

- Estudiar los materiales.
- Revisar un estudio de validación de métodos publicado y criticar la aplicación de la estadística.

Preguntas de Autoevaluación:

- El análisis de regresión lineal para una prueba de hemoglobina glicosilada muestra los siguiente: Nuevo Método (Y) = 0.93 Viejo método (X) + 0.29 ($r = 0.992$). ¿Cuál es el estimado para la pendiente y cuál es el estimado para el intercepto-Y?
- ¿Son confiables los estimados para la pendiente y el intercepto-y?
- ¿Debería usted emplear la prueba-to BlandAltman en lugar de la regresión lineal simple?
- ¿Debería usted usar la regresión de Deming o el modelo Passing-Bablock en lugar de la regresión lineal simple?

Validación de Métodos: Algunos Aspectos a Considerar en el uso de las estadísticas en la validación de métodos

James O. Westgard, PhD

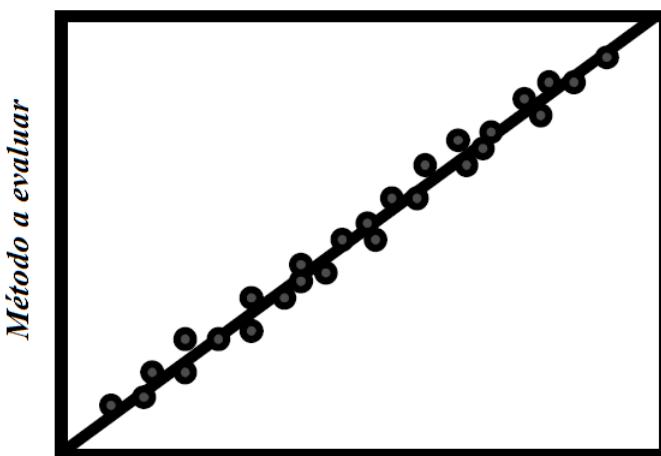
Esta es una actualización de otra versión de una editorial aparecida en Noviembre de 1998 en un ejemplar de la revista "Clinical Chemistry", volumen 4, páginas 2240-2242.

Como químicos clínicos y científicos de laboratorio, a menudo nos preocupamos cuando personal que tiene poca experiencia y entrenamiento en el laboratorio comienza a realizar exámenes de laboratorio, tales como las aplicaciones de *Point of Care*. Si bien hoy en día puede ser fácil realizar esos exámenes con los sistemas analíticos modernos disponibles, aun hay cosas que pueden salir mal. Nosotros esperamos que algún tipo de sistema de la calidad se use para verificar que todo está funcionando correctamente con los análisis de *Point of Care*.

Imagine usted como se sienten los estadísticos considerando que tenemos en nuestras manos programas estadísticos muy poderosos. Estos programas hacen que sea muy simple cargar un conjunto de datos y luego así calcular una gran variedad de estadísticos aunque no sepámonos que representan o para que se emplean. Debemos verificar que en nuestros laboratorios estamos trabajando el análisis estadístico de los datos de manera apropiada.

En un ejemplar de "Clinical Chemistry" aparecido en Noviembre de 1998, Stockl, Dewitte and Thienpont [1], señalaron que la calidad de los datos puede ser más importante que la calidad de la técnica de regresión (es decir, regresión lineal simple versus regresión de Deming versus el modelo Passing-Bablok). En la revista *Clinical Chemistry* el procedimiento estándar para analizar los datos de un experimento de comparación de métodos ha sido preparar un gráfico de comparación que muestre los resultados del método evaluado en el eje "Y" y los resultados del método de comparación en el eje "X" y

Gráfico de Comparación



Método de Comparación

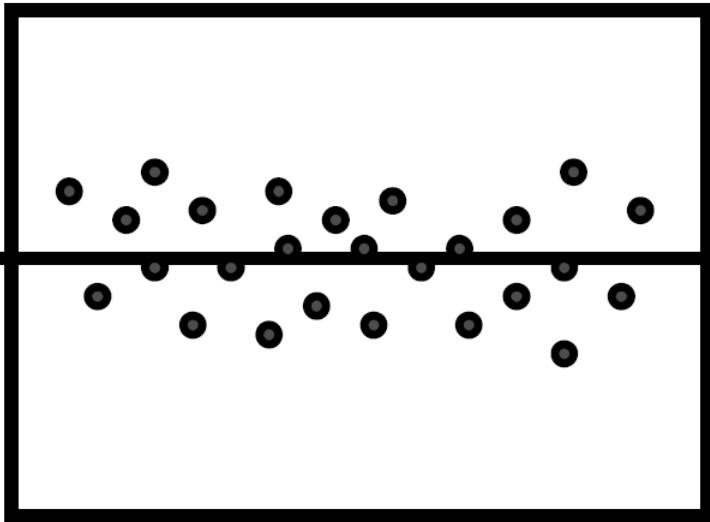
luego calcular las estadísticas de regresión para hallar la mejor línea de ajuste para los datos. Diferentes modelos de regresión pueden ser apropiados, dependiendo de las características de los datos; particularmente el rango analítico que es cubierto considerando los valores del ensayo que son críticos desde un punto de vista clínico.

En varias citas bibliográficas [2], existe una tendencia a desalentar el uso del análisis de regresión y se sugiere su reemplazo por una simple representación gráfica de los datos de la comparación de métodos con el formato de un *gráfico de diferencias* (Difference Plot), en el cual se representa la diferencia entre los valores obtenidos entre el método a evaluar y el de referencia en el eje "Y" versus los valores obtenidos por el método de referencia en el eje "X", en el caso de contar con un método de referencia, o la diferencia entre los valores obtenidos entre el método a evaluar y el de comparación en el eje "Y" versus la media de los valores del método a evaluar y de comparación en el eje "X". Se conoce a este gráfico como Bland-Altman [3]. Hyltoft-Petersen y colaboradores [4] han demostrado que el gráfico de diferencias debe ser cuidadosamente interpretado para tomar una decisión objetiva sobre el desempeño de un método. No es tan simple utilizar este gráfico de diferencias cuando debe realizarse una interpretación objetiva para evaluar el desempeño de un método.

A pesar de estos informes y recomendaciones recientes sobre uso de las estadísticas, muchos analistas e investigadores todavía tienen dificultades con los datos de comparación de métodos. Nosotros hemos estudiado eso problemas hace treinta años [5]. Muchos de los problemas que observamos hace tiempo, son similares a los problemas que existen hoy en día, la diferencia es que actualmente existen programas de computación que simplifican mucho los cálculos. No ha habido mucha mejora, si es que hubo alguna, en los conocimientos y habilidades sobre estadística básica en los laboratorios a la fecha, no solamente para los estudio de validación de métodos sino también para control estadístico

Método a evaluar - Métodos de comparación

Gráfico de Diferencia



Método de Comparación

de la calidad. Esto no significa que no ha habido mejoras en la teoría y que no han sido publicadas recomendaciones en la literatura, más bien implica que las prácticas de laboratorio no han cambiado mucho realmente.

Por lo tanto aun necesitamos prestar mucha atención para juntar, analizar e interpretar los datos de los estudios de comparación de métodos. Aquí presentamos algunos puntos sobre los que debemos prestar cuidado:

Punto N° 1: Usar la estadística para brindar un estimado de los errores, no como un indicador de aceptabilidad. Tal vez este sea el punto principal para tener un sentido práctico sobre el uso de las estadísticas en los estudios de comparación de métodos. Recuerde que el propósito de los estudios de validación es estimar o validar las especificaciones para los parámetros de desempeño del método. [Ver *El significado secreto mas interno, oculto y profundo.*] Las características de aplicación y metodológicas deberían ser consideradas durante el proceso de selección del método. [Ver *Seleccionando un método para validar.*] Las estadísticas son simples herramientas para cuantificar el tamaño de los errores a partir de los datos recolectados durante los experimentos de validación de métodos. [Ver *Conjunto de herramientas para el análisis de datos*].

Las estadísticas no van a decirle a usted directamente si un método es aceptable, si le van a brindar estimados de los errores que le permitirán a usted evaluar la aceptabilidad del método en estudio. Usted hace esto comparando la magnitud del error observado con el tamaño del error que podría ser tolerado sin comprometer la interpretación clínica de los resultados del examen. El desempeño del método no es considerado aceptable cuando la magnitud del error observado es mayor al tamaño del error que podría ser tolerado. El proceso de decisión puede ser facilitado empleando un criterio matemático [6] o a partir del uso de herramientas gráficas [7]. [Ver *La decisión sobre el desempeño del método.*]

Punto N° 2: Reconocer que el propósito principal del experimento de comparación de métodos es obtener un estimado del error sistemático o Sesgo. El experimento de comparación de métodos se realiza para estudiar la exactitud de un nuevo método. [Ver *El plan experimental.*] Es fundamental obtener información sobre el error sistemático promedio o sesgo. También es útil para identificar la naturaleza proporcional o constante del error sistemático y para cuantificar el error aleatorio entre ambos métodos. Conocer la naturaleza del error es importante porque nos orienta a tomar acciones específicas en el laboratorio para controlar el error total del proceso de análisis (por ejemplo, reducir el error sistemático proporcional mejorando la calibración del método). El error total es importante para juzgar sobre la aceptación del método y puede ser calculado a partir de sus componentes (errores sistemáticos y aleatorios).

Punto N° 3: Obtener estimaciones del error sistemático a concentraciones que representen o estén próximas a niveles de decisión médica.

La recolección de especímenes y la selección de las estadísticas pueden ser optimizadas considerando la concentración o concentraciones donde la interpretación de los resultados va a ser mas críticamente interpretada desde un punto de vista médico. [Ver el resumen de *Niveles de decisión médica* por Dr. Bernard Statland.] Si existe una única concentración de decisión médica los datos para la comparación de métodos deben ser obtenidos cerca de ese valor (es decir que no va a ser necesario un amplio rango de datos) y el gráfico de diferencias va a ser de utilidad (junto con la estimación del sesgo a partir de los estadísticos “t”, ver el punto N° 8 más abajo). Si hay dos o más concentraciones que representen niveles de decisión médica, es recomendable obtener especímenes que cubran el rango analítico del método, usar un gráfico de comparación para visualizar los datos, y calcular las estadísticas de regresión para estimar el error sistemático a cada nivel de decisión médica.

Punto N° 4: Cuando existe un único nivel de decisión médica hacer la estimación del error sistemático cerca de la media de los datos. El principal aspecto a tener en cuenta cuando existe un único nivel de decisión médica es juntar los datos alrededor de ese nivel de decisión. La elección de las estadísticas no va a ser crítica cuando hay un único nivel de decisión médica relevante y está próximo a la media de los datos. La estimación del Sesgo a partir de la prueba-t para datos apareados y la estimación del error sistemático a partir de las estadísticas de regresión van a generar la misma estimación del error.

[Nota de explicación: El sesgo es la diferencia entre las medias de dos métodos ($\text{Sesgo} = Y_{\text{promedio}} - X_{\text{promedio}}$), lo cual es equivalente al promedio de las diferencias calculadas de los pares de datos a partir de la prueba-t. Con las estadísticas de regresión, el error sistemático (ES) es estimado a una concentración crítica dada, X_c , como sigue: $ES = Y_c - X_c$, donde Y_c es calculado a partir de las estadísticas de regresión por la ecuación $Y_c = a + b.X_c$, donde “a” es el intercepto-Y y “b” es la pendiente de la línea de regresión. En la regresión lineal simple, la pendiente se calcula en primer lugar, luego se determina el intercepto-Y a partir de la siguiente ecuación: $a = Y_{\text{promedio}} - b.X_{\text{promedio}}$. Cuando en nivel de decisión es igual a X_{promedio} , entonces él $ES = (a + b.X_{\text{promedio}}) - X_{\text{promedio}} = Y_{\text{promedio}} - b.X_{\text{promedio}} + b.X_{\text{promedio}} - X_{\text{promedio}} = Y_{\text{promedio}} - X_{\text{promedio}}$, es decir el mismo estimado del error sistemático será obtenido a partir de las estadísticas de regresión y las estadísticas de la prueba-t, aún cuando el rango de datos evaluados sean pequeño y los valores para pendiente e intercepto-Y no son confiables.]

Punto N° 5: Cuando existen dos o más niveles de decisión médica, utilice el coeficiente de correlación, “r”, para evaluar si el rango de datos es el adecuado para utilizar el modelo de regresión lineal simple. Como fue confirmado por Stockl, Dewitte, y Thienpont [1], cuando “r” es 0.99 o mayor, el rango de datos debería ser lo suficientemente ancho para que el modelo de regresión lineal simple brinde estimados confiables para la pendiente y el intercepto-Y. Resaltan que que si “r” es menor que 0.975, los estadísticos generados a partir de la regresión lineal simple no serán confiables y sugieren mejorar el rango de datos o recurrir a modelos de regresión alternativos más apropiados para esta situación. Note que no se utiliza “r” para evaluar la aceptación del desempeño del método, pero si es utilizado para evaluar la aceptación del rango de datos utilizado para calcular las estadísticas de regresión.

Punto N° 6: Cuando el valor de “r” es alto, utilice el gráfico de comparación junto con las estadísticas de la regresión lineal simple. La confiabilidad de la pendiente y del intercepto-Y son afectadas por datos aberrantes y errores de no linealidad así como también por el rango de concentraciones de los datos [5]. Los datos aberrantes deben ser identificados preferiblemente en el momento graficando de manera inmediata los datos en el gráfico de comparación; los datos discrepantes deben ser investigados mientras los especímenes aún están disponibles. El error por no linealidad puede ser identificado a partir de la visualización del gráfico de comparación, el rango de concentraciones puede ser restringido al segmento lineal, y las estadísticas recalculadas. Stockl, Dewitte, y Thienpont [1] recomiendan usar el gráfico de residuales como parte de análisis de regresión e inspeccionar la secuencia de inicio de los residuales para realizar esta evaluación.

Punto N° 7: Cuando “r” es pequeño, mejore el rango de concentraciones de los datos o cambie el modelo estadístico de regresión. Considere alternativas para mejorar el rango de concentraciones de los datos, reduciendo la variación del método de comparación haciendo replicados, estimando el error sistemático para la media de los datos, dividiendo los datos en subgrupos cuyas medias coincidan con los niveles de decisión médica (los cuales pueden ser luego analizados empleando la prueba-t y el gráfico de diferencias) o empleando técnicas de regresión más sofisticadas. Stockl, Dewitte, y Thienpont [1] hallaron que el modelo de regresión de Deming es más satisfactorio que la técnica de Passing-Bablok. Note que estas técnicas de regresión no están siempre disponibles en los programas ordinarios de estadística aunque si podrá hallarlas en programas espacialmente diseñados para la estudios de validación de métodos de laboratorio.

Punto N° 8: Cuando “r” es bajo y se utiliza el gráfico de diferencias, calcule las estadísticas a partir de la prueba-t para obtener una estimación cuantitativa del error sistemático. Planteado el objetivo principal de estimar el error sistemático a partir de los datos del estudio de comparación de métodos, la utilidad del gráfico de diferencias por si solo es cuestionable ya que la interpretación visual será influenciada principalmente por la dispersión o error aleatorio observado entre los métodos. El sesgo, o promedio de las diferencias a partir de los resultados de los pares de datos debería ser calculado. Los programas de computación para calcular las estadísticas de la prueba-t van a darnos esta estimación junto con el desvío estándar de las diferencias, el cual da una estimación cuantitativa de la dispersión entre los métodos. Note que esa dispersión entre métodos depende de la imprecisión del método evaluado, la imprecisión del método de comparación e interferencias que afectan a los especímenes individuales de distinta manera por los dos métodos. El valor “t” considerado individualmente es un cociente entre el error sistemático y el error aleatorio, y es principalmente útil para determinar si se ha recolectado una cantidad suficiente de datos para hacer confiable a la estimación del sesgo (nuevamente evite usar las estadísticas como un indicador para la aceptabilidad del método). Mientras Bland-Altman también recomiendan el cálculo de la media de las diferencias y la desviación estándar de las diferencias y sugieren que la media de las diferencias mas menos dos desviaciones estándar ($\pm 2S$) sean trazadas en el gráfico [3], es incorrecto juzgar la aceptabilidad de las diferencias observadas por comparación con ellas mismas. Para acceder a una discusión más profunda sobre como juzgar la aceptación de un método sobre la base de los gráficos de diferencias vea Hyltoft-Petersen y colaboradores [4].

Punto N° 9: Cuando existan dudas sobre la validez de la técnica estadística empleada, vea si los resultados obtenidos a partir de las distintas técnicas modifica la decisión sobre la aceptación del desempeño del método. Ya que es muy fácil realizar los cálculos a partir de los programas de estadística disponibles, se pueden comparar como los resultados obtenidos a partir de la distintas técnicas de regresión afectan la interpretación del desempeño del método. Si la selección de la técnica estadística afecta su decisión sobre la aceptabilidad del método, entonces tenga cuidado. Por lo general lo mejor será juntar más datos y asegurarse que la interpretación de esos nuevos datos satisfacen las conclusiones de la técnica de análisis seleccionada.

Punto N° 10: Planifique el experimento cuidadosamente y junte los datos necesarios considerando la técnica estadística que va a usar. Usted puede juntar los datos para ajustarlos a las necesidades del modelo estadístico

empleado o cambiar el modelo estadístico para compensar la limitación de los datos. Una correcta comprensión del uso y aplicación de las estadísticas le va a ayudar a planificar los experimentos y minimizar las dificultades que pudiesen surgir en la interpretación de los resultados. [Ver *El experimento de comparación de métodos*] Si usted está estableciendo un proceso estándar para la validación de métodos en su laboratorio, sería aconsejable hacer un esfuerzo para juntar los datos apropiados, esta es mi recomendación personal para los laboratorios del área de la salud. Este enfoque de obtener los datos correctos implica tomar la muestra apropiada, en las condiciones correctas, procesar las muestras correctamente, operando el método o sistema analítico en condiciones representativas y analizar la muestra del paciente con un proceso que se encuentre bajo control estadístico. Este punto requiere mucho cuidado y debería tener la más alta prioridad. Las estadísticas no tiene sentido si usted no tiene realmente cuidado con los datos.

En términos de gestión de la calidad, el uso correcto de las estadísticas es un problema crónico que continuará generando problemas hasta que sea resuelto. El proceso que necesita ser mejorado es el proceso de educación y entrenamiento en las áreas de química clínica, patología clínica y laboratorio clínico. Existe una deficiencia en una competencia básica – la habilidad de usar la estadística básica para los estudios de validación de métodos y control estadístico de la calidad. Para corregir esta deficiencia se necesitan cursos para estudiantes en formación, talleres de educación continua y seminarios para profesional que ya están en el campo, así como la publicación de artículos en literaturas científicas para recordarle a los investigadores las dificultades y problemas. Con el acceso que en la actualidad nos brinda Internet, cursos básicos de entrenamiento y herramientas de software pueden ser enviadas a quien las quiera donde quiera y cuando quiera.

Referencias

1. Stockl D, Dewitte K, Thienpont M. Validity of linear regression in method comparison studies: limited by the statistical model or the quality of the analytical data? Clin Chem 1998;44:2340-6.
2. Hollis S. Analysis of method comparison studies [editorial]. Ann Clin Biochem 1996;33:1-4.
3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;307-10.
4. Hyltoft-Petersen P, Stockl D, Blaabjerg O, Pedersen B, Birkemose E, Thienpont L, Flensted Lassen J, Kjeldsen J. Graphical interpretation of analytical data from comparison of a field method with a reference method by use of difference plots [opinion]. Clin Chem 1997;43:2039-46.
5. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method-comparison studies. Clin Chem 1973;19:49-57.
6. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974;20:825-33.
7. Westgard JO. A method evaluation decision chart for judging method performance. Clin Lab Science 1995;8:277-83.

Referencias en Línea

Points of care in using statistics in method comparison studies. Trabajo original en formato PDF. <http://www.westgard.com/downloads/>

Use and interpretation of common statistical tests in method-comparison studies. Trabajo original en formato PDF. <http://www.westgard.com/downloads/>

Medical Decision Levels. <http://www.westgard.com/decision.htm>

19: ¿Cómo pueden ser verificadas las especificaciones de desempeño de los fabricantes?

En este capítulo el Dr. Neill Carey, presidente del comité de la CLSI EP 15 que redactó los lineamientos para la guía “Verificación del usuario de las características de desempeño de precisión y veracidad declaradas por el fabricante” colabora con el Dr. Westgard para describir el protocolo y el análisis de datos recomendado a los laboratorios para verificar las especificaciones de desempeño declarados por los fabricantes.

Objetivos:

- Comprender el protocolo experimental de la guía EP 15 para precisión y veracidad.
- Aprender como realizar los cálculos de la EP 15 empleando una planilla de cálculo de Excel.
- Aprender cómo realizar los cálculos para “veracidad” usando la prueba-t para datos apareados con una hoja de cálculo.
- Aplicar los lineamientos de la EP 15 para verificar las especificaciones de desempeño de los fabricantes para precisión y veracidad.

Materiales del capítulo:

- VM-EP 15 Guía para verificar las especificaciones de desempeño declarados por el fabricante para precisión y veracidad, R. Neill Carey, PhD, y James O. Westgard, PhD

Cosas qué hacer:

- Estudiar los materiales.
- Configurar los cálculos de la guía EP 15 para precisión empleando una hoja de cálculo.
- Configurar los cálculos para la prueba-t con datos apareados empleando una hoja de cálculo.
- Verificar los cálculos de su hoja de cálculos empleando el conjunto de datos ejemplo de la guía EP 15.

Preguntas de Autoevaluación:

- ¿Qué es lo que hace al protocolo EP 15 atractivo para los laboratorios?
- ¿Cuáles son las dificultades potenciales en la implementación del EP 15?
- ¿Qué es la precisión intra laboratorio?
- ¿Qué es veracidad?
- ¿Cuál es el modelo estadístico recomendado por el EP 15 para la estimación del sesgo?

Validación de Métodos: Guía para verificar las especificaciones de desempeño declarados por el fabricante para precisión y veracidad

R. Neill Carey, PhD, y James O. Westgard, PhD

Propósito

Hoy en día los laboratorios trabajan con la presión de gestionar a la calidad al menor costo posible. Esta cuestión de costos también alcanza a los estudios de validación de métodos. ¿Cuál es el estudio mínimo necesario para satisfacer el requisito regulatorio de verificar las especificaciones de desempeño declaradas por los fabricantes? Esta cuestión fue abordada directamente en la guía de la CLSI EP 15: “Verificación del usuario de las características de precisión y veracidad declaradas por el fabricante.”[1]

Este documento de CLSI fue desarrollado por un comité formado por usuarios de los laboratorios, profesionales, fabricantes y el gobierno. El proceso de CLSI implica revisión por pares de la guía recomendada por votación oficial de sus miembros. Una vez aprobada, la guía representa “buenas prácticas de laboratorio” que son aceptadas por la comunidad profesional, organismos de acreditación e inspección, así como también por el gobierno. El laboratorio solo necesita referenciar el EP 15 para justificar el uso de este protocolo para satisfacer las regulaciones de CLIA para la validación de nuevos métodos.

De acuerdo al subcomité de la EP 15 el propósito de este documento es el siguiente:

“El subcomité tuvo dos metas principales durante el desarrollo de la EP 15. Una fue desarrollar un protocolo de examen que fuera lo suficientemente simple para ser aplicado en laboratorios de una gran variedad de recursos y complejidad, desde los pequeños laboratorios ubicados en las salas de los patólogos hasta los grandes laboratorios clínicos. La segunda meta fue desarrollar un protocolo lo suficientemente riguroso como para brindar conclusiones estadísticamente válidas para los estudios de verificación. Para cumplir con esas dos metas el subcomité desarrolló un protocolo de examen de cinco días y hojas de trabajo simplificadas para juntar todos los datos, realizar cálculos estadísticos y trabajar las pruebas para evaluar la precisión y veracidad observadas. Se brinda una hoja de cálculo para simplificar y estandarizar los cálculos estadísticos y pruebas para la precisión y veracidad observadas.”

“Este documento fue diseñado para ser usado primariamente cuando un método ya establecido (validado por el fabricante) es inicialmente puesto en funcionamiento en el laboratorio. También puede ser empleado para verificar el desempeño de un método después de la implementación de una acción correctiva posterior a una falla en una encuesta de un esquema de evaluación de la competencia.”

Alcance y Definiciones

El EP 15 ha sido diseñado para que los laboratorios verifiquen que sus métodos tienen un desempeño comparable al desempeño declarado por el fabricante en sus especificaciones. Otros documentos de CLSI ofrecen lineamientos a los fabricantes para establecer sus características de desempeño para precisión (EP 5) [2] y veracidad (EP 9) [3].

El EP 15 brinda un protocolo y análisis de datos para verificar las especificaciones de desempeño del fabricante para repetibilidad, precisión intra laboratorio y veracidad. Estas características de desempeño están definidas en el EP 15 de la siguiente forma:

- **Repetibilidad (de los resultados de una medición):** Proximidad o acuerdo entre los resultados de mediciones sucesivas del mismo mensurando llevadas a cabo bajo las mismas condiciones de medición. NOTA: Formalmente se empleaba el término de precisión intra corrida o intra serie.
- **Condiciones de Repetibilidad:** Condiciones a partir de las cuales resultados independientes de un ensayo son obtenidos con el mismo método sobre el mismo material en el mismo laboratorio por el mismo operador usando el mismo equipo en un período corto de tiempo.
- **Precisión Intra laboratorio:** Precisión obtenida en un período determinado de tiempo; operadores, calibración y lotes de reactivos pueden variar dentro del mismo laboratorio y usando el mismo equipo. NOTA: Formalmente se utilizaba el término de precisión total.
- **Veracidad (de una medición):** Proximidad o acuerdo entre el valor promedio obtenido a partir de una serie de resultados obtenidos sobre el mismo material y un valor de referencia aceptado. NOTA: La medida de la veracidad se expresa usualmente en términos de sesgo.

El EP 15 describe tres protocolos: uno para precisión basado en el análisis de tres replicados sobre un material de control durante cinco días; otro para veracidad basado en el análisis de 20 especímenes de pacientes; y un tercer protocolo de veracidad basado en el análisis, por lo menos, de dos materiales de referencia con valores verdaderos asignados.

EP 15 Protocolo de Precisión

Se especifica un período de familiarización para aprender a operar el sistema analítico, incluyendo calibración, procedimientos de mantenimiento y procedimientos de monitoreo (control estadístico interno de la calidad). Los procedimientos de control de la calidad recomendados por el fabricante van a ser usados para monitorear el desempeño del sistema analítico durante el desarrollo del protocolo EP 15. Los materiales de control a utilizar en el protocolo deben ser seleccionados considerando que sus concentraciones deben estar próximas o representar niveles de decisión médica. Las concentraciones de los materiales de control también deben estar próximas a las concentraciones empleadas por los fabricantes al momento de establecer sus especificaciones para precisión. Si es posible, deberían ser los mismo materiales que empleó el fabricante al momento de establecer sus especificaciones, o muy similares (matriz similar). El protocolo de examen [1] es el siguiente:

1. “Analizar una corrida por día con tres replicados por muestra para cada una de las dos (o tres) concentraciones seleccionadas a diario durante cinco días.
2. Si una corrida debe ser rechazada a causa de los resultados del procedimiento de control interno de la calidad o dificultades operativas, descartar los datos y agregar una corrida adicional.
3. Incluir dentro de la corrida analítica del protocolo los controles diarios para monitorear el desempeño del método durante el protocolo.
4. Las muestras para veracidad deben ser incluidas dentro de la misma corrida.
5. Calibrar como especifica el fabricante en sus instrucciones para la operación rutinaria método en el equipo. Si el fabricante menciona en sus especificaciones de desempeño que los datos para precisión fueron obtenidos considerando ciclos múltiples de calibración, entonces el operador puede elegir calibrar también durante el experimento.”

EP 15 Ecuaciones para los Cálculos de Precisión

S intra corrida (s_r), Varianza para las medias diarias (s_b^2), S intra laboratorio (s_i), y grados de libertad efectivos (T).

$$s_r = \sqrt{\frac{\sum_{d=1}^D \sum_{i=1}^n (x_{di} - \bar{x}_d)^2}{D(n-1)}}$$

Ecuación 1

$$s_b^2 = \frac{\sum_{d=1}^D (\bar{x}_d - \bar{\bar{x}})^2}{D-1}$$

Ecuación 2

$$s_i = \sqrt{\frac{n-1}{n} \cdot s_r^2 + s_b^2}$$

Ecuación 3

$$T = \frac{((n-1) \cdot s_r^2 + (n \cdot s_b^2))^2}{\left(\frac{n-1}{D}\right) \cdot s_r^4 + \left(\frac{n^2 \cdot (s_b^2)^2}{D-1}\right)}$$

Ecuación 4

Dónde D Número total de días,

N número total de replicados por día,

x_{di} resultado del replicado “i” del día “d”,

\bar{x}_d media de los resultados del día “d”,

$\bar{\bar{x}}$ media de todos los resultados.

Cálculo de Datos. El EP 15 brinda estimaciones para repetibilidad (S_r) y precisión intra laboratorio (S_i). Las ecuaciones son algo complicadas, como puede verse en la página 236. Los cálculos implican determinar las varianzas intra corrida y entre corridas, luego combinar las dos para conseguir la varianza total, la cual es la base para calcular el desvío estándar intra laboratorio (S_i). Recuerde que la varianza es el cuadrado de la desviación estándar, por lo tanto la precisión intra laboratorio es la raíz cuadrada de la varianza total del laboratorio. Hay una hoja de trabajo en el EP 15 que lo orienta paso a paso en estos cálculos y además hay un ejemplo demostrando todos estos cálculos.

El EP 15 hace referencia a una hoja de cálculo comercial para realizar estos cálculos, pero la CLSI aun no ha lanzado esta herramienta de cálculo hasta comienzos del 2008. Por esta razón, abajo damos algunos lineamientos para realizar cálculos. Si usted tiene acceso a una herramienta de cálculo para EP 15, estos lineamientos no le serán necesarios y usted solamente necesita comprender los resultados y como son los mismos interpretados para verificar las especificaciones de desempeño de los fabricantes.

Cálculos empleando Excel. Un ejemplo a partir de un conjunto de datos simulados para EP 15 aparece en la próxima figura, la cual también describe como fueron realizados esos cálculos empleando una hoja de cálculo de Excel. Solo con el propósito de simplificar los cálculos es conveniente ingresar los tres replicados en una columna para cada uno de los cinco días. La media y la desviación estándar (S) para cada día puede ser calculada para cada día usando las funciones PROMEDIO y DESVEST disponibles en Excel. La última columna (G) muestra la varianza de la corrida, obtenida al elevar al cuadrado el S de la

	A	B	C	D	E	F	G
1	EP 15 Precisión						
2	Numero de Replicados	3					
3	Numero de Días	5					
4		Rep. 1	Rep. 2	Rep. 3	Media de la corrida	DE de la corrida	Varianza
5	Día 1	140	140	140	140.000	0.000	0.000
6	Día 2	138	139	138	138.333	0.577	0.333
7	Día 3	143	144	144	143.667	0.577	0.333
8	Día 4	143	143	142	142.667	0.577	0.333
9	Día 5	142	143	141	142.000	1.000	1.000
10							
11	Media promedio	=promedio(E5:E9)			141.333		
12	Varianza Intra corrida (Vr)	=promedio(G5:G9)			0.400		
13	S Intra corrida	=raiz(E12)			0.632		
14	S Entre corrida	=desvest(E5:E9)			2.147		
15	Varianza Entre corrida (Vb)	=E14*E14			4.611		
16	Cociente Vr/Vb	+E12/E15			0.087		
17	Varianza Total	((B2-1)/B2)*E12+E15			4.878		
18	S del Laboratorio	=raiz(E17)			2.209		
19	CV del Laboratorio	(E18/E11)*100			1.563		
20							

corrida presente en la columna anterior (F).

Con la información de la figura de arriba, las estimaciones de repetibilidad y precisión intra laboratorio pueden ser calculadas siguiendo los siguientes pasos:

- Obtener la media promedio usando la función de Excel PROMEDIO para los datos comprendidos entre las celdas E5 y E9 (141.333 mg/dL para este ejemplo de EP 15 de glucosa).
- Obtener la varianza intra corrida promedio empleando la función PROMEDIO de Excel para los datos comprendidos entre las celdas G5 y G9. (0.400 mg²/dL² para este ejemplo)
- Obtener la raíz cuadrada de la varianza intra corrida promedio de la celda E12 aplicando la función de Excel RAIZ. (0.632 mg/dL, que es el desvío estándar intra corrida, es decir en condiciones de repetibilidad)
- Calcular el desvío estándar entre corridas S usando la función de Excel DESVEST para los datos comprendidos entre las celdas E5 y E9. (2.147 mg/dL para este ejemplo)
- Elevar al cuadrado el valor obtenido en la celda E14 empleando la función de Excel POTENCIA y así obtener la varianza entre corridas. (4.611 mg²/dL² para este ejemplo)
- Calcular el radio entre la varianza intra corrida (V_r) celda E12 y la varianza entre corridas (V_b) celda E15.(0.087 para este ejemplo)
- Calcular la varianza total partiendo de las varianzas intra corrida (V_r) y la varianza entre corridas(V_b); Nota: Esto se hace multiplicando la varianza intra corrida (V_r de la celda E12) por un factor (n-1)/n (dónde n es el número de replicados que se ingresó en la celda B2) y luego sumando el valor de la varianza entre corridas (V_b de la celda EP 15).(4.878 mg²/dL² para este ejemplo)
- Calcular el desvío estándar S intra laboratorio aplicando la función de Excel RAIZ al valor obtenido en la celda E17.(2.209 mg/dL que es el desvío estándar intra laboratorio)
- Calcular el coeficiente de variación intra laboratorio CV dividiendo el valor del desvío estándar intra laboratorio obtenido en la celda E18 por la media obtenida en la celda E11, y multiplicando al número resultante de ese cociente por 100. (1.563 % para este ejemplo)

El ejemplo de esta hoja de cálculo fue configurado para un diseño de tres replicados durante 5 días, no obstante más replicados pueden ser añadidos

insertando columnas después de la Repetición (Rep.) 3, columna D y más días pueden ser agregados insertando filas después del Día 5 (fila 9). La inserción de columnas y filas debe ser hecha de esta forma para asegurar que las ecuaciones aplicadas a los datos siguen siendo válidas.

Verificación de las especificaciones de precisión: El EP 15 recomienda que el “valor de verificación” sea calculado para generar un límite superior para la especificación declarada por el fabricante para precisión. Si el desvío estándar S observado por el laboratorio es menor que el valor de verificación, entonces la especificación del fabricante ha sido verificada. El valor de verificación depende del número de replicados (n), del número de días (D), de la varianza intra corrida (S_r^2 , en condiciones de repetibilidad) y de la varianza entre corridas (S_b^2 , varianza para las medias diarias). El procedimiento consiste en calcular un parámetro llamado “T”, el cual representa los “grados de libertad efectivos”, luego se busca en una tabla de doble entrada “CHI CUADRADO” en la fila correspondiente al valor de “T” y en la columna correspondiente a la cantidad de niveles de control empleados, el valor de “C”, luego se calcula el valor de verificación para la especificación declarada por el fabricante ($S_{declarada}$) para precisión de la siguiente forma:

$$\text{Valor de Verificación} = S_{\text{declarada}} * C^{1/2}/T^{1/2}$$

El primer paso es calcular “T” de acuerdo a la ecuación 4 (mostrada en un gráfico anterior). Esta ecuación es confusa y difícil de implementar en una única celda en una hoja de cálculo de Excel. Es por esto que se muestra un cálculo paso a paso en la hoja de cálculo de Excel de la siguiente página:

	A	B	C	D	E
1	Calculo de T				
2					
3	Numero de Replicados	3	3	3	3
4	Numero de Días	5	5	5	5
5	Varianza Intra corrida (Vr)	0.40	0.00	4.00	12.00
6	Varianza Entre corrida (Vb)	4.61	4.00	4.00	4.00
7	Radio cociente entre varianzas Vr/Vb	0.09	0.00	1.00	3.00
8	(n-1)Vr	0.80	0.00	8.00	24.00
9	nVb	13.83	12.00	12.00	12.00
10	Suma de los dos anteriores	14.63	12.00	20.00	36.00
11	Cuadrado del número de arriba = Numerador	214.12	144.00	400.00	1296.00
12	(n-1)/D	0.4	0.40	0.40	0.40
13	(n-1/D)Vr ²	0.06	0.00	6.40	57.60
14	n ² (Vb ²)/(D-1)	47.84	36.00	36.00	36.00
15	Suma de los dos anteriores = Denominador	47.9	36.00	42.40	93.60
16	T calculado del cociente Numerador/Denominador	4.47	4.00	9.43	13.85

Esta hoja de cálculo muestra el cálculo de “T” para el EP 15 a partir de un conjunto de datos ejemplo en la columna B, además ilustra la gama de valores de “T” que pueden ser obtenidos dependiendo del tamaño relativo de las varianzas intra corrida y entre corridas (B16, C16, D16, E16). Note que para simplificar la interpretación en la hoja de cálculo, Vr representa la varianza intra corrida (S_r^2) y Vb representa la varianza entre corridas (S_b^2). El EP 15 para el conjunto de datos de la columna B da un valor de 4.47. Note que el rango de “T” puede variar desde D-1, o 4, a n*D-1, o 14, como se muestra en los cálculos adicionales en las columnas C, D y E. Si la varianza intra corrida es pequeña comprada con la varianza entre corridas, entonces “T” será bajo. Un valor bajo de “T” puede representar a un método automatizado que tiene una muy buena precisión intra corrida (repetibilidad), pero con variables de operación que causan desvíos día a día lo que queda evidenciado en la varianza entre corridas. Un valor alto de “T” puede representar a un instrumento de pie de cama que usa cartuchos descartables para cada medición. La varianza intra corrida puede ser grande comprada con la varianza entre corridas, en este caso el valor de “T” será grande. Si la varianza entre corridas es cero “0”, “T” da un valor de 10. Como “T” puede variar tanto, será necesario ir sobre los cálculos confusos de arriba.

Una vez que el valor de “T” ha sido calculado, el valor de “C” puede ser hallado en la tabla de “Puntos seleccionados de porcentaje para una distribución Chi cuadrado para diferente cantidad de niveles con el fin de brindar un porcentaje de falso rechazo del 5 %”, la cual puede ser encontrada al final de este capítulo. Para el conjunto de datos ejemplo del EP 15 el valor de “T” es 4.47 y se asume que se trabajó con dos niveles de control. Si entrámas en la tabla en la columna de 4 grados de libertad efectivos (fila 4) y en la columna que representa 2 niveles de control (columna B) el valor de “C” es 11.14.

Con estos valores de “C” y “T”, podemos estimar el “valor de verificación” para la especificación declarada por el fabricante ($S_{declarada}$) de la siguiente forma:

$$\text{Valor de Verificación} = S_{declarada} * C^{1/2}/T^{1/2}$$

Para nuestro ejemplo de EP 15 el fabricante declaró un S intra laboratorio de 2.00 mg/dL. Dado un valor de “C” de 11.14 y siendo “T” 4.47, sus raíces cuadradas son 3.34 y 2.11 respectivamente.

$$\text{Valor de Verificación} = 2.00 * 3.34/2.11 = 3.16 \text{ mg/dL}$$

La estimación del laboratorio para la precisión intra laboratorio es de 2.21 mg/dL, que es menor que 3.16 mg/dL, esto significa que los datos del laboratorio verifican la especificación declarada por el fabricante para precisión.

Si bien el EP 15 recomienda verificar precisión en condiciones de repetibilidad y precisión intra laboratorio, será suficiente verificar únicamente la precisión intra laboratorio ya que es más representativa para las condiciones de operación de rutina de los laboratorios.

EP 15 Protocolo para Veracidad empleando Muestras de Pacientes

Es una práctica común para validar el desempeño de un nuevo método analizar muestras de pacientes por el nuevo método y el método que está siendo reemplazado. Se asume que los resultados del método de comparación (viejo) son correctos y esta es la base sobre la cual los fabricantes establecen sus especificaciones de desempeño para veracidad o sesgo. Es probable que esto sea correcto cuando el nuevo método es una actualización del método previo o del sistema analítico anterior. Si esto no es así la especificación del fabricante debe ser cuidadosamente inspeccionada para estar seguros de que el método de comparación en el laboratorio y el empleado por el fabricante al momento de establecer sus especificaciones son comparables en desempeño.

El siguiente es el protocolo recomendado:

1. Analizar 20 muestras cuyas concentraciones cubran el rango reportable del método evaluado;
2. Analizar muestras frescas en las mismas condiciones de operación de rutina en el laboratorio;
3. Medir de 5 a 7 muestras por día durante un período de 3 o 4 días por el método a evaluar y el método de comparación no dejando pasar más de 4 horas en el procesamiento por uno y otro método;
4. Evaluar los resultados del control estadístico interno de la calidad para asegurar que el sistema analítico está trabajando de manera estable y validar los resultados obtenidos;
5. Inspeccionar los datos comparados para identificar resultados discrepantes;
6. Calcular las diferencias entre los pares de datos y graficar esas diferencias versus los valores obtenidos por el método de comparación para obtener una visualización gráfica de los datos;
7. Someter a los pares de datos a una prueba-t de datos apareados para determinar la media de las diferencias entre métodos (sesgo) así como también la desviación estándar de las diferencias (S_{dif});

8. Calcular los límites de confianza y/o verificación para comparar el sesgo observado con el declarado por el fabricante en sus especificaciones de desempeño.

Cálculo de Datos

Los cálculos de la prueba-t para datos apareados son soportados por varios programas de estadística. El “calculador de datos apareados” descripto en el Capítulo 7 realizará estos cálculos y preparará un gráfico de diferencias con los resultados (Ver <http://www.westgard.com/mvtools.html>).

Cálculos usando una hoja de cálculo de Excel. Es fácil configurar estos cálculos en una hoja de cálculo de Excel, como se muestra a continuación:

	A	B	C	D	E
1	EP15 Veracidad				
2		Resultado Método a evaluar	Resultado Método comparación	Y_i-X_i	(Y_i-X_i-B)
3	1	76	77	-1.00	-3.50
4	2	127	121	6.00	3.50
5	3	256	262	-6.00	-8.50
6	4	303	294	9.00	6.50
7	5	29	25	4.00	1.50
8	6	345	348	-3.00	-5.50
9	7	42	41	1.00	-1.50
10	8	154	154	0.00	-2.50
11	9	398	388	10.00	7.50
12	10	93	92	1.00	-1.50
13	11	240	239	1.00	-1.50
14	12	72	69	3.00	0.50
15	13	312	308	4.00	1.50
16	14	99	101	-2.00	-4.50
17	15	375	375	0.00	-2.50
18	16	168	162	6.00	3.50
19	17	59	54	5.00	2.50
20	18	183	185	-2.00	-4.50
21	19	213	204	9.00	6.50
22	20	436	431	5.00	2.50
23	Suma	3980	3930	50.00	0.00
24		199.00	196.50	2.50	4.33
25		Media (Y)	Media (X)	Sesgo	DE_{Dif}
26		=promedio(B3:B22)	=promedio(C3:C22)	=promedio(D3:D22)	=desvest(E3:E22)
27	Intervalo de Confianza	t-crítico	n	superior	inferior
28	p=0.01	2.861	20	5.27	-0.27
29	Límite de Verificación	Especificación de Fabricante		superior	inferior
30	p=0.01	2		4.77	-0.77

La columna A puede ser empleada para colocar las identificaciones de las muestras, los resultados del método a evaluar se ingresan en la columna B y los resultados del método de comparación se ingresan en la columna C. Las diferencias entre los resultados del método evaluado y el método de comparación se muestran en la columna D. Los promedios de los resultados del método evaluado y del método de comparación se muestran al final de las columnas (199.0 para el método evaluado y 196.5 para el método de comparación). La diferencia entre esos dos promedios nos da un estimado del sesgo, el cual es de 2.50 mg/dL. El mismo estimado puede ser obtenido calculando el promedio de las diferencias obtenidas en la columna D. La columna E muestra la diferencia individual de cada par de datos menos el sesgo entre los métodos ($Y_i - X_i - B$). La desviación estándar de esa columna brinda un estimado del desvío estándar de las diferencias S_{dif} .

Se puede calcular un intervalo de confianza para el Sesgo observado de 2.50 mg/dL de la siguiente forma:

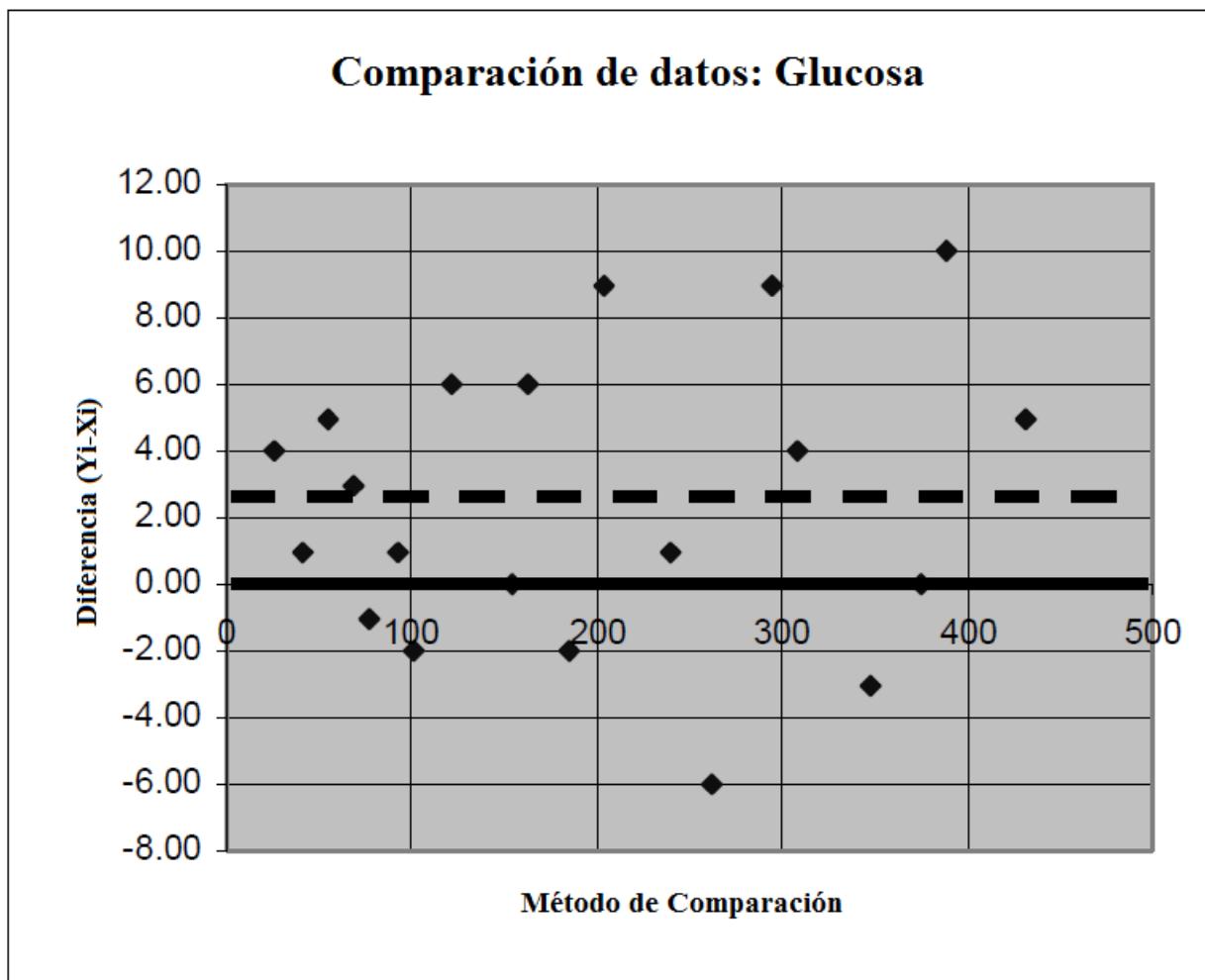
$$\text{Límite superior} = \text{Sesgo} + t_{\text{crítico}} * S_{\text{dif.}} / (N^{1/2})$$

$$\text{Límite inferior} = \text{Sesgo} - t_{\text{crítico}} * S_{\text{dif.}} / (N^{1/2})$$

dónde el sesgo calculado a partir de los datos del ejemplo es de 2.50 mg/dL, $t_{\text{crítico}}$ para N-1 o 19 grados de libertad y un $p=0.01$ es de 2.861, como podemos encontrar en la tabla-t al final de este capítulo, S_{dif} ha sido calcula con un valor de 4.33 mg/dL, y N es de 20 para el protocolo EP 15. El límite superior es de 5.27 mg/dL y el límite inferior es de -0.27 mg/dL, como podemos ver en la hoja de cálculo.

Si la especificación de desempeño del fabricante para el Sesgo es de 2.00 mg/dL, los límites de verificación para la especificación del fabricante pueden ser calculados de la misma manera, lo cual dará unos límites de verificación de 4.77 mg/dL a -0.77 mg/dL como podemos ver en la hoja de cálculo. Dado que el sesgo observado de 2.50 mg/dL cae dentro de los límites de verificación de la especificación declarada por el fabricante, los datos verifican dicha especificación. Note que los límites de confianza se superponen a la especificación del fabricante, por lo tanto esto también verifica que los datos son consistentes con las especificaciones del fabricante.

También es posible emplear las funciones de Excel para efectuar gráficos y preparar así el gráfico de diferencias como vemos en la figura abajo:



En este gráfico la línea sólida representa “cero” diferencias y la línea punteada representa el sesgo observado de 2.50 mg/dL.

EP 15 Protocolo para Veracidad empleando Materiales de Referencia

El EP 15 brinda una segunda aproximación para verificar la veracidad de un método-el análisis de materiales de referencia con valores asignados. Estos materiales deben ser cuidadosamente seleccionados considerando su conveniencia de uso en métodos de análisis clínicos y sistemas analíticos. Los materiales recomendados incluyen los siguientes:

- Materiales certificados de referencia (CRM, por sus siglas en Inglés) disponibles para algunos analitos en E.U.A. en el NIST “National Institute of Standards and Technology” (Instituto Nacional de Estándares y Tecnología);
- Materiales de referencia pertenecientes a esquemas de evaluación de la competencia con valores asignados;
- Materiales provistos por fabricantes con valores asignados;
- Materiales empleados en programas de evaluación externa de la calidad;
- Materiales provistos por terceras partes que tienen valores asignados a través de su análisis en distintos laboratorios;
- Materiales estándar que pueden ser preparados a concentraciones conocidas.

El EP 15 recomienda que se seleccionen por lo menos dos materiales que representen concentraciones próximas a un niveles de decisión médica bajo y a un nivel de decisión médica del rango reportable del método. Los materiales deberían ser preparados de acuerdo a las recomendaciones del fabricante, analizados por duplicado durante 3 o 5 corridas, luego se debe calcular media y desvío estándar junto con los límites de confianza para ayudar a la verificación del valor asignado.

Cálculo de Datos

Un conjunto de datos provistos a manera de ejemplo están disponibles en el EP 15 y se muestran en la hoja de cálculo de Excel abajo:

	A	B	C
1	Material de Referencia con Valor Asignado		40.0 mg/dL
2			
3	Día 1	Replicado 1	37
4		Replicado 2	38
5	Día 2	Replicado 1	39
6		Replicado 2	37
7	Día 3	Replicado 1	38
8		Replicado 2	36
9	Día 4	Replicado 1	39
10		Replicado 2	38
11	Día 5	Replicado 1	38
12		Replicado 2	37
13			
14	Recuento	=contar(C3:C12)	10
15	Media	=promedio(C3:C12)	37.7
16	S	=desvest(C3:C12)	0.95
17	S de la media	C16/raiz(C14)	0.3
18	t-crítico		3.25
19	Límite de Confianza Superior	C15+C18*C17	38.68
20	Límite de Confianza Inferior	C15-C18*C17	36.73
21	DE del Valor Asignado		0.15
22	Incertidumbre combinada	=raiz((potencia(C17))+(potencia(C21)))	0.33
23	Límite de Verificación Superior	C15+C18*C22	38.79
24	Límite de Verificación Inferior	C15-C18*C22	36.61

La media y el desvío estándar se calculan con las funciones de Excel PROMEDIO y DESVEST, luego el error estándar de la media se calcula teniendo en cuenta las 10 mediciones efectuadas. El valor-t crítico para N-1 grados de libertad y una p=0.010 es 3.25, el cual puede ser hallado en la tabla-t al final de este capítulo. El intervalo de confianza del 95% calculado va de 36.7 a 38.7 mg/dL, este intervalo no incluye al valor asignado de 40.0 mg/dL; por lo tanto la veracidad no ha sido demostrada a partir de estos datos experimentales.

El EP 15 recomienda considerar la incertidumbre del valor verdadero asignado combinando la varianza (S_a^2) con la incertidumbre de la medición del experimento realizado. Esto se logra sumando las dos varianzas y luego aplicando la RAIZ cuadrada a las varianzas combinadas. Si bien los límites de verificación van a ser algo más grandes que los límites de confianza, los límites

de confianza serán por lo general apropiados para evaluar si el desempeño observado demuestra la veracidad del método. En el caso del ejemplo del EP 15, la incertidumbre del valor verdadero asignado S_a tiene un valor de 0.15 mg/dL. Cuando S_a se combina con la incertidumbre experimental, el S resultante es de 0.33 mg/dL. Por lo tanto los límites de verificación son ligeramente más amplios que los límites de confianza-36.61 a 38.79 comprados con 36.73 a 38.68- y nuevamente la interpretación es que la veracidad no ha sido demostrada a partir de los datos experimentales.

¿Cuál es el punto?

El muy probable que el protocolo EP 15 se convierta en una práctica estándar en E.U.A. para verificar las especificaciones de precisión y veracidad declaradas por los fabricantes. Para que este protocolo brinde una verificación confiable de las especificaciones declaradas por los fabricantes, los cálculos deben ser llevados a cabo de manera correcta. Los cálculos para precisión son más complejos que los necesarios para el experimento de replicación simple para el cuál se toma una única medición del material por cada día. El período más corto de tiempo propuesto por el EP 15 implica mediciones repetidas de los materiales cada día, esto hace necesario combinar adecuadamente la varianza intra corrida con la varianza entre corridas para obtener una estimación confiable del desvío estándar intra laboratorio. El protocolo de veracidad a partir de muestras de pacientes es el más tradicional, empleando estadísticas de la prueba-t para datos apareados y un gráfico de diferencias. Sin embargo, el protocolo que utiliza materiales de referencia con valores asignados va a necesitar un análisis cuidadoso y la comprensión del nuevo concepto de incertidumbre de la medición. Esto será un paso en la dirección de la comprensión de los conceptos ISO de veracidad e incertidumbre y las prácticas globales emergentes para caracterizar el desempeño de un procedimiento de medida.

Referencias

1. CLSI EP15-A2(e). User Verification of Performance for Precision and Trueness: Approved Guideline – Second Edition. Clinical and Laboratory Standards Institute, Wayne PA, 2005.
2. CLSI EP5-A2. Method Comparison and Bias Estimation using Patient Samples. Clinical and Laboratory Standards Institute, Wayne PA, 2002.
3. CLSI EP9-A2. Evaluation of Precision Performance of Quantitative Measurement Methods. Clinical and Laboratory Standards Institute, Wayne PA, 2004.

Puntos de porcentaje seleccionados para la distribución Chi-cuadrado para un número seleccionado de niveles para dar una tasa de falso rechazo del 5%. (Tabla 1 del EP 15 A2 (e)).

A	B	C	D
Distribución Chi-Cuadrado p=0.05			
gl	2 niveles	3 niveles	4 niveles
3	9.35	10.24	10.86
4	11.14	12.09	10.86
5	12.83	13.84	14.54
6	14.45	15.51	16.24
7	16.01	17.12	17.88
8	17.53	18.68	19.48
9	19.02	20.12	21.03
10	20.48	21.71	22.56
11	21.92	23.18	24.06
12	23.34	24.63	25.53
13	24.74	26.06	26.98
14	26.12	27.48	28.42
15	27.49	28.88	29.84
16	28.85	30.27	31.25
17	30.19	31.64	32.64
18	31.53	33.01	34.03
19	32.85	34.36	35.4
20	34.17	35.7	36.76
21	35.48	37.04	38.11
22	36.78	38.37	39.46
23	38.08	39.68	40.79
24	39.36	41	42.12
25	40.65	42.3	43.35

Valores críticos t para Probabilidades Seccionadas (p) y Grados de Libertad Seleccionados (gl)

A	B	C	D
Tabla - T	Intervalos o pruebas para dos colas		
gl	p=0.10	p=0.05	p=0.01
2	2.92	4.3	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.6
5	2.02	2.57	4.03
6	1.94	2.45	3.71
7	1.9	2.36	3.5
8	1.86	2.31	3.36
9	1.83	2.26	3.25
10	1.81	2.23	3.17
12	1.78	2.18	3.06
14	1.76	2.14	2.98
16	1.75	2.12	2.92
18	1.73	2.1	2.88
20	1.72	2.09	2.84
30	1.7	2.04	2.75
40	1.68	2.02	2.7
60	1.68	2	2.66
120	1.66	1.98	2.62
∞	1.64	1.95	2.58

20: ¿Cómo pueden ser evaluadas las especificaciones de desempeño declaradas por los fabricantes en una escala Sigma?

Las regulaciones CLIA requieren que los laboratorios verifiquen que el desempeño observado para un método es consistente con las especificaciones de desempeño declaradas por el fabricante. Además los laboratorios deberían estar bien asesorados para evaluar la calidad de los procesos de análisis sobre la base de esas especificaciones de desempeño así como también sobre la aceptabilidad de estas especificaciones para el uso clínico previsto para el método. Se puede calcular una métrica Sigma a partir de las especificaciones de desempeño declaradas por el fabricante para precisión (S, CV) y exactitud (Sesgo, Veracidad) para brinda una evaluación objetiva de la calidad del método en una escala Sigma y así juzgar objetivamente sobre la aceptación del mismo.

Objetivos:

- Comprender el concepto Seis Sigma.
- Aprender como calcular una métrica Sigma a partir de la imprecisión, el sesgo y el requisito de la calidad definido para el método.
- Establecer un vínculo entre la métrica Sigma y la aceptabilidad del desempeño del método.

Materiales del capítulo:

- VM-Traduciendo las especificaciones de desempeño en una métrica Sigma, por James O Westgard, PhD, y Sten Westgard, MS

Cosas qué hacer:

- Estudiar los materiales.
- Revisar las aplicaciones de las métricas Sigma en el sitio web de Westgard

Preguntas de Autoevaluación:

- ¿Cuál es el vínculo que existe entre el Gráfico de Decisión de Método y la métrica Sigma?
- ¿Cuál es la ecuación para estimar la métrica Sigma?
- ¿Cómo puede estimar la métrica sigma a partir de las especificaciones de desempeño declaradas por el fabricante?
- ¿Cuál es el desempeño Sigma de un método de colesterol para el cual la precisión declarada por el fabricante es de 2.0 % a una concentración de 200 mg/dL y la exactitud declarada está dada por la siguiente ecuación de regresión $Y=0.96X+6.0$?

Validación de Métodos: Transformando las especificaciones de desempeño en una métrica Sigma

James O. Westgard, PhD y Sten Westgard, MS

El Gráfico de Decisión de Métodos que fue descripto en el capítulo 16 brinda una herramienta gráfica para juzgar sobre el desempeño de un método. El punto operativo representa la precisión y el sesgo observado para el método, el cual es comprado con varios criterios para juzgar sobre la aceptación del desempeño del mismo. Todos esos criterios diferentes describen distintos niveles de la calidad sobre la base de una “escala Sigma”, por ejemplo un método que satisface el criterio “ $TE = \text{Sesgo} + 6 \text{ DE}$ ” alcanza el nivel de la calidad 6 sigma, el cual es el objetivo máximo de la calidad de la clase mundial “world class”, un método que satisface el criterio “ $TE = \text{Sesgo} + 3 \text{ DE}$ ” alcanza la mínima calidad requerida para un proceso de producción.

Las especificaciones de desempeño declaradas por el fabricante para precisión y sesgo pueden ser convertidas directamente en una métrica Sigma para determinar la aceptabilidad del método-para responder a la pregunta, “¿Es este método lo suficientemente bueno?”

Una introducción terriblemente corta a Seis Sigma

Para poder comprender la relación entre los datos de validación de métodos y Seis Sigma, nosotros necesitamos hacer una introducción a los conceptos y cálculos de Seis Sigma. Esto representa un pequeño problema ya que por lo menos esto es un asunto de un libro entero. De hecho nosotros hemos escrito un libro completo sobre el tema: Seis Sigma CC Diseño y Control [1]. Por lo tanto todo lo que nosotros podemos ofrecer en este pequeño espacio, próximo al final del libro, es un perfil resumido de Seis Sigma. Debería ser suficiente para ayudarlo a usted, pero tenga en mente que elementos y conceptos más profundos sobre Seis Sigma que los que nosotros podemos discutir acá.

Seis Sigma es uno de los sistemas de gestión de la calidad más populares. Pero los conceptos de Seis Sigma están disponibles desde hace mucho tiempo. Las métricas Sigma son realmente la evolución de la Gestión Total de la Calidad con una evaluación más cuantitativa del desempeño de los procesos y metas claras para la mejora de la calidad.

Los principios de Seis Sigma se remontan al enfoque que Motorola le aplicó a la Gestión de la Calidad total antes del cambio de siglo. Motorola estableció una meta: **6 Sigmas o desviaciones estándar de la variación propia del procesos debería ajustarse (no superar) a los límites de tolerancia del proceso**, de aquí bien el nombre Seis Sigma.

El poder de Seis Sigma reside en permitir una medición universal del desempeño de un proceso en una “escala Sigma” para facilitar una evaluación comparativa en la industria. La metodología Seis Sigma puede ser aplicada siempre que las “salidas” de un proceso puedan ser medidas. Para muchos procesos las “salidas pobres” pueden ser contadas como errores o defectos, expresadas como defectos por millón de oportunidades (DPM), y luego convertidas en una métrica Sigma empleando una tabla estandarizada disponible en cualquier texto de Seis Sigma [1]. En este tiempo dónde las “salidas” de los procesos son de gran interés para el ámbito de la salud, Seis Sigma brinda una metodología para describir desempeño de una manera cuantitativa en términos que son fáciles de comprender.

Dos modelos para medir el desempeño de un proceso

La Gestión de la Calidad Seis Sigma brinda una meta general para el desempeño de los procesos -“Seis Sigmas” de la variación de un proceso deberían ajustarse dentro de los límites de tolerancia o requisitos de la calidad del producto. Dada esta meta, es importante medir el desempeño del proceso para determinar si el mismo necesita ser mejorado. Hay dos metodologías distintas. Una se basa el conteo de los defectos producidos por un proceso (medida de las “salidas”), y la otra basada en medir la variabilidad del proceso de manera directa (una medición predictiva).

La primera alternativa (inspeccionar las “salidas”) es ampliamente aplicada en la industria y en el ámbito de la salud y depende del conteo del número de defectos a la salida del proceso, estimar la tasa de defectos por millón de oportunidades (DPM), y luego convertir DPM a una métrica Sigma usando una tabla de conversión estándar. Cada vez que estimemos una tasa de error, nosotros podemos usar esta metodología para convertir la tasa de error en una métrica Sigma. Esta primera metodología es fácilmente aplicable a procesos de laboratorio pre y pos analíticos.

La primera aplicación de la métrica Sigma a datos de laboratorio fue publicada en el 2000 por Nevalainen y colaboradores [1]. Los autores utilizaron datos de tres laboratorios individuales, así como también los resúmenes de desempeño obtenidos a partir de entre 300 a 500 laboratorios del programa Q-Probe de CAP (College of American Pathologists). El trabajo original brinda información sobre el tamaño de muestra empleado y el número de defectos o

errores hallados, los cuales fueron convertidos a errores porcentuales y defectos por millón de oportunidades. Los valores en DPM para indicadores de la calidad representativos tomados del Q-Probe son convertidos a la métrica Sigma en la tabla de abajo.

Indicador de la Calidad del Esquema Q-Probe	% de Error	DPM	Sigma*
Exactitud de la solicitud médica	1.8 %	18,000	3.60
Pruebas pedidas por duplicado	1.52	15,200	3.65
Error de identificación de brazalete	0.65	6,500	4.00
Demoras en el informe de TDM	24.4	244,000	2.20
Aceptabilidad de especímenes en hematología	0.38	3,800	4.15
Aceptabilidad de especímenes en química clínica	0.30	3,000	4.25
Registro de muestras de patología durante las cirugías	3.4	34,000	3.30
Adecuación de los especímenes de citología	7.32	73,700	2.95
Esquemas de evaluación de la competencia del laboratorio	0.9	9,000	3.85
Discordancia en el diagnóstico de las muestras congeladas de patología	1.7	17,000	3.60
Frotis de PAP de tamizaje falsos negativo	2.4	24,000	3.45
Informe de errores	0.0477	477	4.80

* Conversión empleando una tabla con una tolerancia de desvío de 1.5DE.

Este modelo que mide “salidas” puede ser virtualmente aplicado a cualquier proceso. Varios de los errores observados para estos procesos están en un rango de 0.3 % a 3.0 %, lo que se traduce típicamente a una métrica Sigma en un rango de 4 a 3. El desempeño analítico, estimado a partir de la participación en los esquemas de evaluación de la competencia tiene apenas un sigma de 3.85. El mejor proceso es “Informe de errores” con una métrica Sigma de 4.80. Ninguno de los procesos evaluados cumple con la meta del desempeño Seis Sigma. La métrica de desempeño sigma promedio de todos estos procesos es de 3.65.

Solo a efectos de hacer una evaluación comparativa, Nevalainen cito los siguientes números. El manejo de equipaje de una aerolínea muestra una tasa de error de 0.4 %, o 4000 DPM, lo cual representa un 4.15 considerando la métrica Sigma. La seguridad en las aerolíneas (del sistema normal de causas aleatorias, no causas asignables, tales como secuestros terroristas) tiene un índice de fatalidad muy bajo de 0.43 muertes por millón de pasajeros por milla, lo cual es mejor aún que la meta de desempeño de 6 Sigmas para una calidad de clase mundial “world class”. En la industria, el desempeño mínimo aceptable para un proceso de producción se establece por lo general en 3.0 Sigmas. Por lo general se espera que los procesos de negocios tengan un desempeño de alred-

edor de 4.0 Sigmas. Como fue observado por Nevalaine, los procesos típicos del laboratorio tienen un desempeño sigma similar al desempeño promedio de los procesos de negocios. Solo para considerar un ejemplo diferente, la producción de neumáticos de Firestone-los neumáticos de las SUV que reventaron causaron cerca de 2000 accidentes de auto con 100 muertes-tenía un desempeño de 5 Sigmas, que es mejor que cualquiera de los desempeños de los procesos de laboratorio evaluados por Nevalaine.

Los hallazgos de Nevalaine no fueron una aberración; la mayoría de los procesos de laboratorio no están siquiera cerca de tener un desempeño Seis Sigma. Usted puede considerar esto muy fuerte, pero tasas de error de 1% a 2% son consideradas excelentes para procesos en el ámbito de la salud. Esas tasas de error corresponden a 10,000 y 20,000 DPM respectivamente o un nivel de desempeño Sigma de 3.8 a 3.6. Actualmente deberíamos estar apuntando a tasas de error de 0,1 % (4.6 Sigmas) a 0,01% (5.2 Sigmas) y por ultimo a 0,001 % (5.8 Sigmas).

Es difícil emplear esta metodología de conteo de defectos para determinar el desempeño Sigma para los exámenes de laboratorio. El problema reside en establecer que es un error en un resultado de laboratorio: ¿Es el resultado un error cuando difiere del valor verdadero de lo que se pretende medir? (y ¿cómo determina usted cual es el valor verdadero de lo que pretende medir en la muestra del paciente si usted la mide solamente una vez?) o ¿Es el resultado un error cuando el desempeño del examen excede los 2 DE o cuando excede algún otro límite del control o excede el error clínico tolerable? o ¿es el resultado un defecto si causa un error en la decisión médica, o si el resultado causa alguna fatalidad?. La mayoría de los laboratorios, si emplean como es usual como límite para sus controles 2 DE, estarían determinando erróneamente el conteo de defectos y por ende obtendrían un valor incorrecto para la métrica Sigma del proceso analítico que están evaluando. Considerando estas dificultades, no vamos a usar este modelo con los datos provenientes de la validación de métodos.

La segunda metodología (medir variación) es actualmente mucho más simple y se vincula de una manera natural con los datos provenientes de la validación de métodos. Esta metodología depende de definir los requisitos de la calidad y la variación del procedimiento de medida. Ya que esto es lo que se hace exactamente durante los estudios de validación, toda la estadística y datos obtenidos durante la validación del método pueden ser inmediatamente utilizados para el cálculo de la métrica Sigma.

Evaluación de la calidad de los procesos analíticos

Es muy fácil acceder al desempeño de un procedimiento de medida analítico en una escala Sigma. Los niveles máximos de tolerancia (Requisitos de la Calidad) pueden ser tomados a partir de los criterios que CLIA aplica a sus esquemas de evaluación de la competencia o también pueden ser seleccionados a partir de alguna otra fuente adecuada; la variación del proceso y el sesgo pueden ser estimados a partir de los experimentos de validación de métodos, datos de esquemas de comparación de grupo par (peer group “Esquemas Interlaboratorio”), resultados de esquemas de evaluación externa de la calidad o evaluación de la competencia, o inclusive datos obtenidos del control estadístico interno de la calidad de rutina.

Esto es lo que usted necesita de un estudio de validación de métodos:

- Resultados de los estudios de replicación (CV).
- Resultados de los estudios de comparación de métodos (pendiente, intercepto-Y e inclusive r).

Esto es lo que usted necesita a partir de otras fuentes:

- Requisitos de la Calidad (CLIA, Clínicos, Biológicos u otra fuente).
- Niveles de Decisión Médica (para determinar concentraciones críticas).

Para calcular la métrica Sigma, debe tomar el requisito de la calidad seleccionado, restarle el sesgo observado para su método y dividir al número resultante por el DE o CV de su método como se ve en la siguiente ecuación:

$$\text{Sigma del Método} = \frac{\text{(Requisito de la Calidad} - \text{Sesgo del Método)}}{\text{CV del Método}}$$

El primer paso para estimar el Sigma del Método es obtener las estimaciones de Sesgo y precisión (CV) del/los nivel/es de decisión médica críticos escogido/s (X_c). Por lo general los fabricantes ofrecen especificaciones de desempeño para precisión estimadas a 2 o 3 concentraciones diferentes. Si alguna de estas concentraciones esta próxima al nivel de decisión de interés, entonces usar el DE o CV obtenido por el fabricante para esa concentración. En algunos casos es necesario promediar los coeficientes de variación si el nivel de decisión médica de interés está entre las concentraciones de los materiales evaluados por los fabricantes; otras veces puede ser apropiado interpolar entre los coeficientes de variación declarados por los fabricantes para las concentraciones que el evaluó. Es mejor trabajar con coeficientes de variación (CV) porque el desvío estándar cambia notoriamente con las concentraciones y el coeficiente de variación no

tanto. Además la mayoría de los requisitos de la calidad provistos por CLIA para desempeño aceptable están expresados en porcentaje.

Los fabricantes presentan por lo general las especificaciones de desempeño para sesgo con el formato de la ecuación de una línea recta obtenida a partir de una regresión, es decir, $Y_c = a + bX_c$, donde X_c representa el nivel de decisión de interés, Y_c es la mejor estimación para el valor observado para el nuevo método, y “a” y “b” son el intercepto-Y y la pendiente que han sido calculados a partir de los datos del experimento de comparación de métodos llevado a cabo por el fabricante. Calcular Y_c , luego tomar la diferencia entre Y_c y X_c como una estimación del sesgo, luego dividir por X_c y multiplicar por 100 para tener el sesgo expresado en porcentaje. Note que todos los componentes de la ecuación de la métrica Sigma deben estar expresados en las mismas unidades (Requisito de la Calidad, CV y Sesgo), ya sea unidades de concentración o porcentaje.

Ejemplo de Cálculo de la Métrica Sigma

Tomemos un ejemplo para Colesterol. Para este ejemplo el fabricante ha establecido sus especificaciones a partir de un experimento en el cual dos niveles de control han sido analizados una vez por día durante 30 días:

Especificación para precisión total:

1.9 % para una concentración de 106.7 mg/dL

1.6 % para una concentración de 238.7 mg/dL

Para exactitud el fabricante ha realizado un estudio de comparación de métodos que incluye 80 especímenes de pacientes que han sido analizados por el nuevo método y el sistema analítico de la generación anterior (como método de comparación).

Especificación para Sesgo

$$Y = 1.46 + 0.99 X$$

Requisito de la Calidad

CLIA: 10 %

Ya que los niveles de decisión médica importante para colesterol son de 200 y 240 mg/dL, comenzaremos con el nivel de decisión de 240 mg/dL. Para estimar el sesgo a 240 mg/dL, $Y_c = 1.46 + 0.99 (240) = 1.46 + 237.6 = 239.1$ mg/dL. El sesgo es la diferencia observada entre 239.1 y 240 mg/dL o sea 0.9 mg/dL (en valor absoluto), y expresada en porcentaje, $((0.9/240) \times 100)$, 0.4 %. Ya que la imprecisión para 238.7 mg/dL es de 1.6 % y la concentración evaluada por el

fabricante (238.7) es muy próxima al nivel de decisión médica evaluado (240), podemos emplear 1.6 % como estimado de CV.

Por lo tanto el sigma será de: (10 % - 0.4 %) / 1.6 %; es un procedimiento de medida 6.0 Sigmas al nivel de decisión médica de 240 mg/dL.

Para estimar el sesgo a 200 mg/dL, $Y_c = 1.46 + 0.99 (200) = 1.46 + 198 = 199.5$ mg/dL. El sesgo es la diferencia observada entre 199.5 y 200 mg/dL o sea 0.5 mg/dL (en valor absoluto), y expresada en porcentaje, ((0.5/200) X 100), 0.3 %. La precisión debería estar entre 1.9 % y 1.6 % más cerca de 1.6 %, por lo tanto tomaremos un valor de 1.7 % como nuestro estimado.

Por lo tanto el sigma será de: (10 % - 0.3 %) / 1.7 %; es un procedimiento de medida 5.7 Sigmas al nivel de decisión médica de 200 mg/dL.

En este caso las métricas Sigmas observadas para ambos niveles de decisión médica son prácticamente iguales, como era de esperarse ya que los niveles de decisión médica evaluados (200 y 240 mg/dL) no están muy separados entre sí.

Algunos lineamientos generales para el cálculo de la métrica sigma

1. Identifique las especificaciones de desempeño declaradas por el fabricante para precisión y sesgo; preste mucha atención al identificar las concentraciones a las cuales las especificaciones fueron obtenidas.
2. Defina los niveles de decisión médica de interés para usted y su laboratorio.
3. Defina el requisito de la calidad para cada uno de los niveles de decisión de interés.
4. Use el CV el coeficiente de variación declarado por el fabricante obtenido a la concentración más próxima al nivel de interés, o promedie los CVs si el nivel de decisión está entre las concentraciones para las cuales el fabricante estableció sus especificaciones, o interpole entre los CVs declarados por el fabricante si su nivel de decisión de interés está próximo a uno de los CVs enmarcados por el fabricante. Para estar más seguro también puede usar el CV más alto declarado por el fabricante.
5. Estime el sesgo al/los nivel/es de decisión de interés.
6. Calcule la métrica Sigma.
7. Si ha trabajado con dos o más niveles de decisión diferentes, compare las métricas sigma para determinar cuál es más crítica para el cuidado de la salud del paciente.

El criterio científico va a jugar un rol importante al momento de seleccionar los números correctos para incluir en los cálculos. Los fabricante por lo general establecen especificaciones para la precisión intra corrida (repetibilidad) y para precisión total (intra laboratorio). El uso de los CVs obtenidos para precisión total nos va a dar Sigmas más bajos, pero esta situación será más representativa del desempeño en el laboratorio a largo plazo. Los fabricantes pueden brindar especificaciones para sesgo obtenidas por comparación con dos o más métodos. Usted debería seleccionar el método de comparación más parecido a su propio método.

El poder de la métrica Sigma

La métrica Sigma establece con un solo número el vínculo entre el requisito de la calidad, la precisión y el sesgo de un método. Establece una escala para el desempeño de un método independientemente de que método estemos hablando. Esta escala es una escala de calidad que ha sido bien establecida en la industria a través de estudios de evaluación comparativa- 6.0 Sigmas es la meta, 3.0 Sigmas es el mínimo desempeño aceptable.

Usted puede determinar facialmente el desempeño que es deseable para un método. Por ejemplo, si el criterio para desempeño aceptable de CLIA para un método es del 10 %, el método debe tener un CV de 1.7 % y un sesgo de 0.0 % para lograr un desempeño sigma de 6.0 $[(10.0-0.0)/1.7]=6.0$. Uno se pregunta cómo un grupo de expertos como los de (NCEP) “*National Cholesterol Education Program*” aparece con guías que establecen un CV máximo aceptable del 3.0 % y un sesgo máximo aceptable del 3.0 % para el colesterol. Si consideramos el requisito de la calidad de CLIA para colesterol del 10 %, un método de colesterol que cumple con esta guía tendrá un desempeño sigma de 2.33 $[(10-3)/3]$. En cualquier otro tipo de industria, un proceso con un desempeño sigma menor de 3.0 no sería considerado para uso de rutina o de producción. Con este ejemplo, usted puede ver como los conceptos y principios de Seis Sigma pueden impactar sobre la visión que tenemos del desempeño de un método y así cambiar nuestras metas y especificaciones para los métodos analíticos.

Evaluación de la calidad para los procesos vinculados a la salud

Dada la universalidad de la aplicación de esta metodología Seis Sigma, podemos esperar que el público en algún momento nos demande saber cuál es la calidad de los procesos vinculados a la salud en una escala sigma. Si no hacemos esta traducción, otros tomaran las tasas de error y las convertirán en una métrica Sigma. El público en general pronto tendrá una mejor compresión de la calidad vinculada a los procesos de la salud!

¿Le gustaría a usted explicarle a un paciente por que los procesos vinculados a la salud son peores que los procesos de manejo de equipaje de las aerolíneas? Probablemente no. Aún no se ha comprendido como evaluar a la calidad de nuestros procesos y como establecer metas para la mejora de estos procesos. Seis Sigma cambia esto. Los laboratorios son afortunados porque estos conceptos pueden ser fácilmente aplicados y los datos necesarios para efectuar los cálculos ya están disponibles.

Este capítulo aborda someramente la superficie de los conceptos, ventajas y aplicaciones de Seis Sigma. Las métricas Sigma son una extensión lógica de la validación de métodos y hacen un uso práctico de los datos de la validación. Los cálculos son simples y los resultados poderosos.

Para obtener más información, consulte *Six Sigma QC Design and Control*, o visite la web de Westgard para acceder a listados de aplicaciones y lecciones sobre Seis Sigma: <http://www.westgard.com/archives.htm>

Referencias

1. Westgard JO. Six Sigma Quality Design and Control: Desirable precision and requisite QC for laboratory testing processes. Second Edition. Madison, WI:Westgard QC, Inc., 2007.
2. Nevalainen D, Berte L, Kraft C, Leigh E, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. Arch Pathol Lab Med 2000;124:516-519.

21: ¿Qué impacto tendrá ISO en la gestión de la calidad analítica?

Este capítulo considera un enfoque general sobre la gestión de la calidad analítica de acuerdo a las recomendaciones de ISO “*International Standards Organization*” volcadas en la norma 15189 para laboratorios clínicos. Teniendo en cuenta guías de Metrología para laboratorios, ISO enfatiza los conceptos de veracidad e incertidumbre. Es muy importante estar al tanto de los conceptos de esta norma, aunque también debemos considerar las fortalezas y debilidades de este modelo en comparación a los modelos tradicionales que se basan en un esquema de errores.

Objetivos:

- Revisar los Requisitos de ISO para la validación y verificación de métodos.
- Identificar y comprender los conceptos de ISO de veracidad, trazabilidad y estimación de la incertidumbre de la medición.
- Comparar estos conceptos metrológicos con los conceptos de error que han sido diseñados para ser empleados en el laboratorio clínico.
- Evaluar el valor práctico que aporta la utilización del modelo de error en la gestión de la calidad analítica en el entorno del laboratorio clínico.
- Identificar cual será el impacto que ISO va a tener sobre los laboratorios de E.U.A. en el futuro.

Materiales del capítulo:

- **Conceptos sobre la calidad – ¿Es mejor estar con incertidumbre que con error?, por James O Westgard, PhD**

Cosas qué hacer:

- Estudiar los materiales.
- Revisar los puntos 5.5 y 5.6 de la Norma ISO 15189 si la tiene disponible.

Preguntas de Autoevaluación:

- ¿Cuál es el significado de las siglas ISO?
- ¿A qué concepto se aproxima más “veracidad”: precisión, exactitud o error total?
- ¿A qué concepto se aproxima más “incertidumbre”: precisión, exactitud o error total?
- ¿Por qué es importante la trazabilidad?
- ¿Cuáles son las ventajas que aportan los conceptos de veracidad e incertidumbre?
- ¿Cuáles son las ventajas de aplicar el modelo de error tradicional en el laboratorio clínico?

Conceptos sobre la Calidad - ¿Es mejor estar con incertidumbre que con error?

James O. Westgard, PhD y Sten Westgard, MS

A partir de la publicación de la Norma ISO 15189 en el año 2003, hay disponible un estándar general de la calidad para ser aplicado en los laboratorios clínicos [1]. Anteriormente algunos estándares de ISO fueron aplicados en el entorno del laboratorio clínico aunque no eran específicos sino de aplicación general, por ejemplo la serie ISO 9000 para “Sistemas de Gestión de la Calidad”, o aplicables para laboratorios de Metrología, por ejemplo la Norma ISO 17025 “Requisitos Generales para la Competencia de Laboratorios de Prueba y Calibración”, la cual enfatiza los conceptos de veracidad e incertidumbre de la medición [GUM, 2].

Con la aparición de la ISO 15189, los conceptos y terminologías que se aplicaban a los laboratorios de prueba y calibración, es decir conceptos metrológicos, se aplicaron a los laboratorios clínicos. Ahora nosotros vivimos en un mundo de veracidad, exactitud, precisión e incertidumbre de medida. Los nuevos documentos de CLSI demuestran que esta terminología está siendo aplicada en los laboratorios de E.U.A., por ejemplo la guía EP15-A2 esta titulada como “Verificación del usuario de las características de desempeño de Precisión y Veracidad” [3]. Adicionalmente la CLSI tiene un proyecto en desarrollo (C51) para generar un documento sobre “Expresión de la Incertidumbre de la Medición en el Laboratorio Clínico”.

Lineamientos de la Norma ISO 15189 para la Validación de Métodos y el Control de la Calidad

La Norma ISO 15189 es un estándar de alto nivel que brinda lineamientos generales sobre lo que debe ser hecho, pero no especifica mucho como hacerlo. Los lineamientos para la validación de métodos y control de la calidad se encuentran en el apartado 5. Las citas que acá realizamos provienen de un documento borrador del 2007 - 2da edición que está próximo a ser editado pero que al momento de publicar este libro no es aún oficial. Si bien pueden aparecer algunos cambios en la edición final de la Norma ISO 15189:2007, la intención es brindarle la mejor información disponible al momento de publicar el libro. Se aconseja a los lectores consulten la versión final del documento que será publicada supuestamente para fines del 2008.

5.5.1.2 Validación y/o Verificación de procedimientos de examen.

Todos los procedimientos de examen deben ser validados para asegurar, brindando evidencia objetiva, que las especificaciones de desempeño del procedimiento

se vinculan y son aptas para el uso previsto del procedimiento de examen. Para aquellos procedimientos que han sido validados por el desarrollador del método (por ejemplo, el fabricante o autor de un método publicado), el laboratorio debe obtener información del desarrollador del método para confirmar que las especificaciones de desempeño del método son apropiadas para el uso previsto del mismo.

Nota 1. La información también puede ser empleada, si es necesario, para validar el método para un uso diferente al previsto por el desarrollador.

Nota 2. Las especificaciones de desempeño del método de examen incluyen: Límite de detección; límite de cuantificación; linealidad, sensibilidad; medición de la precisión, incluyendo medición en condiciones de repetibilidad y en condiciones de reproducibilidad; selectividad/especificidad, incluyendo sustancias interferentes y robustez.

Nota 3. El laboratorio debe tener en cuenta la meta deseada para la medición de la incertidumbre considerando el uso previsto del procedimiento de examen.

5.5.1.3 Validación de procedimientos de examen.

La validación deberá ser tan profunda como sea necesario para confirmar, brindando evidencia objetiva, que los requisitos (especificaciones de desempeño) para el uso previsto específico del procedimiento de examen son cumplidos.

A continuación vemos un listado de la información que debe ser registrada:

Identidad del mensurando, propósito del procedimiento de examen y metas para su desempeño, principio del método, requisitos de desempeño (ver la Nota 2 arriba), tipo primario de muestra, requisitos de equipo y materiales, procedimientos de calibración, instrucciones paso a paso para llevar a cabo el procedimiento de examen, procedimientos de control estadístico interno de la calidad, interferencias, cálculos, medición de la incertidumbre, intervalos de referencia, rango reportable, valores críticos y de alerta, interpretación de la prueba, precauciones de seguridad y fuentes potenciales de variación.

Para cada estudio de validación, el laboratorio debe registrar lo siguiente:

Criterios de aceptación y rechazo, resultados obtenidos, procedimientos de calibración y control, análisis de datos, características de desempeño determinadas, comparación de los resultados con otros métodos y otros laboratorios, factores que afectan los resultados, arrastre (cuando sea aplicable), e interferencias o problemas de especificidad.

5.5.1.3 Verificación de procedimientos de examen.

Los procedimientos de examen de los desarrolladores que son empleados sin modificaciones deben ser verificados para asegurar que cumplen con los requisitos necesarios de los usuarios de acuerdo a su uso previsto. El proceso de verificación debe confirmar, brindando evidencia objetiva, que los requisitos específicos de desempeño son cumplidos. El proceso de verificación debe ser tan profundo como lo requiera el uso previsto del procedimiento de examen. Las especificaciones de desempeño estudiadas durante el proceso de verificación deben ser aquellas que sean relevantes al uso previsto del mismo. El laboratorio debe verificar la instalación del o los equipos y operación del procedimiento de examen.

Para métodos cuantitativos,

se debe verificar que los siguientes parámetros de desempeño tengan un comportamiento semejante al obtenido por el fabricante al momento de su validación: medición de la precisión, incluyendo precisión en condiciones de repetibilidad y reproducibilidad; acuerdo a los resultados de los pacientes con respecto a métodos previos o de referencia; recuperación del mensurando a lo largo del rango de medición analítico; arrastre para altas concentraciones cuando sea aplicable; interferencias y perdidas de especificidad; intervalos de referencia y puntos de corte (puntos de decisión clínica).

Para métodos cualitativos,

se deben verificar los siguientes parámetros de desempeño: precisión en condiciones de reproducibilidad y repetibilidad para muestras por encima y por debajo del valor de corte (positivas y negativas); acuerdo para los resultados de los pacientes (sensibilidad y especificidad diagnóstica); arrastre, cuando sea aplicable; interferencias y problemas de especificidad.

5.6.2.1 Programas de Control de la Calidad.

El laboratorio debe diseñar un programa de control de la calidad para cada procedimiento de examen con la capacidad de detectar errores, monitorear el desempeño del procedimiento de examen y verificar la calidad pretendida para los resultados.

5.6.3 Medición de la Incertidumbre.

Los laboratorios deben determinar la incertidumbre de medida de sus procedimientos de medida con el propósito de interpretar los valores medidos. Cuando se calcula la incertidumbre de medida, todos los componentes de la incertidumbre que contribuyen de manera significativa a la incertidumbre de la medición deben ser tenidos en cuenta empleando un método apropiado de análisis. El laboratorio debe tener los datos de incertidumbre de la medición

de sus procedimientos de medida cuando le sea requerida por los usuarios de sus servicios.

5.6.3.1 Calibración de los sistemas de medida.

El laboratorio debe tener un programa para calibrar los sistemas de medición que está diseñado y realizado para asegurar que los resultados tienen trazabilidad a un material de referencia específico o procedimiento de medición de referencia de alta jerarquía, preferiblemente al SI de unidades o por referencia a una constante natural u otra referencia establecida. Cuando nada de esto sea posible o pertinente, se deben aplicar otros medios para proporcionar confianza en los resultados, incluyendo al menos los siguientes:

- a) participación en un programa adecuado de comparaciones interlaboratorios;
- b) uso de materiales de referencia adecuados, certificados para indicar la caracterización del material;
- c) análisis o calibración por otro procedimiento;
- d) mediciones por cociente o basadas en relaciones de reciprocidad;
- e) normas o métodos de consentimiento mutuo que estén claramente establecidos, especificados, caracterizados y mutuamente acordados por todas las partes involucradas;
- f) documentación de declaraciones respecto de reactivos, procedimientos o el sistema de análisis cuando la trazabilidad la proporciona el proveedor o el fabricante.

5.6.4 Verificando la comparabilidad metrológica de los resultados.

Para aquellos análisis realizados utilizando distintos procedimientos, equipamientos, o en lugares diferentes, o todos éstos, debe existir un mecanismo definido para verificar la comparabilidad de los resultados a lo largo de intervalos clínicamente apropiados. Tal verificación se debe realizar en período de tiempos definidos y apropiados a las características del procedimiento o el instrumento.

Uso previsto, Validación y Verificación

Los documentos ISO emplean una nomenclatura estandarizada que puede no resultar completamente familiar a los científicos del laboratorio. Por ejemplo, la diferenciación entre los conceptos de validación y verificación tiene que ver con el uso previsto. Uso previsto, un término ampliamente empleado en las guías ISO, no está explícitamente definido, pero está vinculado a la calidad requerida para la aplicación clínica prevista de la prueba de laboratorio. El término Uso previsto puede estar representado por un error máximo permitido, tal como se emplea el concepto en este libro. No obstante, también puede estar representado

por un intervalo de decisión clínica, que es un tipo de requisito de la calidad (ver capítulo 2) que es una representación más concreta de la calidad clínica requerida.

Nótese que el concepto de uso previsto aplica al diseño del procedimiento de control estadístico interno de la calidad, específicamente cuando se menciona verificar la calidad prevista de los resultados. Lineamientos sobre como diseñar un esquema de control estadístico interno de la calidad sobre la base de la calidad requerida para una prueba de laboratorio y la precisión y exactitud observada para un método pueden ser encontrados en nuestros libros *Diseño y Control de la Calidad Seis Sigma*, y *Asegurando la Calidad Correcta de la Manera Correcta*.

De acuerdo a la Norma ISO 15189 se define **validación** como *la confirmación, brindando evidencia objetiva, de que los requisitos específicos para un uso previsto u aplicación han sido cumplidos*. Por otro lado, se define **verificación** como *la confirmación, brindando evidencia objetiva, de que requisitos específicos han sido cumplidos*. En el contexto de este libro, cuando hablamos de verificación nos referimos a la confirmación de las especificaciones de desempeño declaradas por el fabricante, como fue discutido en el capítulo 19. Cuando hablamos de validación nos referimos a la comparación de las características de desempeño de un método con un requisito de la calidad establecido para el mismo, lo cual constituye el mensaje principal de este libro, como se demuestra en el gráfico de decisiones en el capítulo 16 y en el de evaluación de las métricas seis sigma de las especificaciones declaradas por el fabricante en el capítulo 20.

Características de Desempeño del Método

El listado de características de desempeño de ISO es muy similar al de CLIA, por ejemplo, rango reportable o linealidad; precisión en condiciones de repetibilidad y reproducibilidad; exactitud o sesgo o veracidad; interferencia, recuperación, y especificidad analítica; límite de detección o límite de cuantificación; intervalos de referencia además de puntos de corte o decisión clínica. ISO enfatiza sobre el concepto de calibración, particularmente en la verificación de la trazabilidad en lugar de demostrar simplemente que los calibradores son consistentes respecto a su valor asignado (lo cual puede implicar o no trazabilidad respecto al valor verdadero). La otra gran diferencia es el énfasis que pone ISO en la incertidumbre de la medición, una característica de desempeño separada que los laboratorios deben determinar.

Si bien este libro brinda una guía detallada sobre cómo definir el Uso Previsto de una prueba y sobre como satisfacer los requisitos de ISO para validación y verificación de las características de desempeño de un método, la discusión es insuficiente para generar una comprensión de los conceptos de veracidad

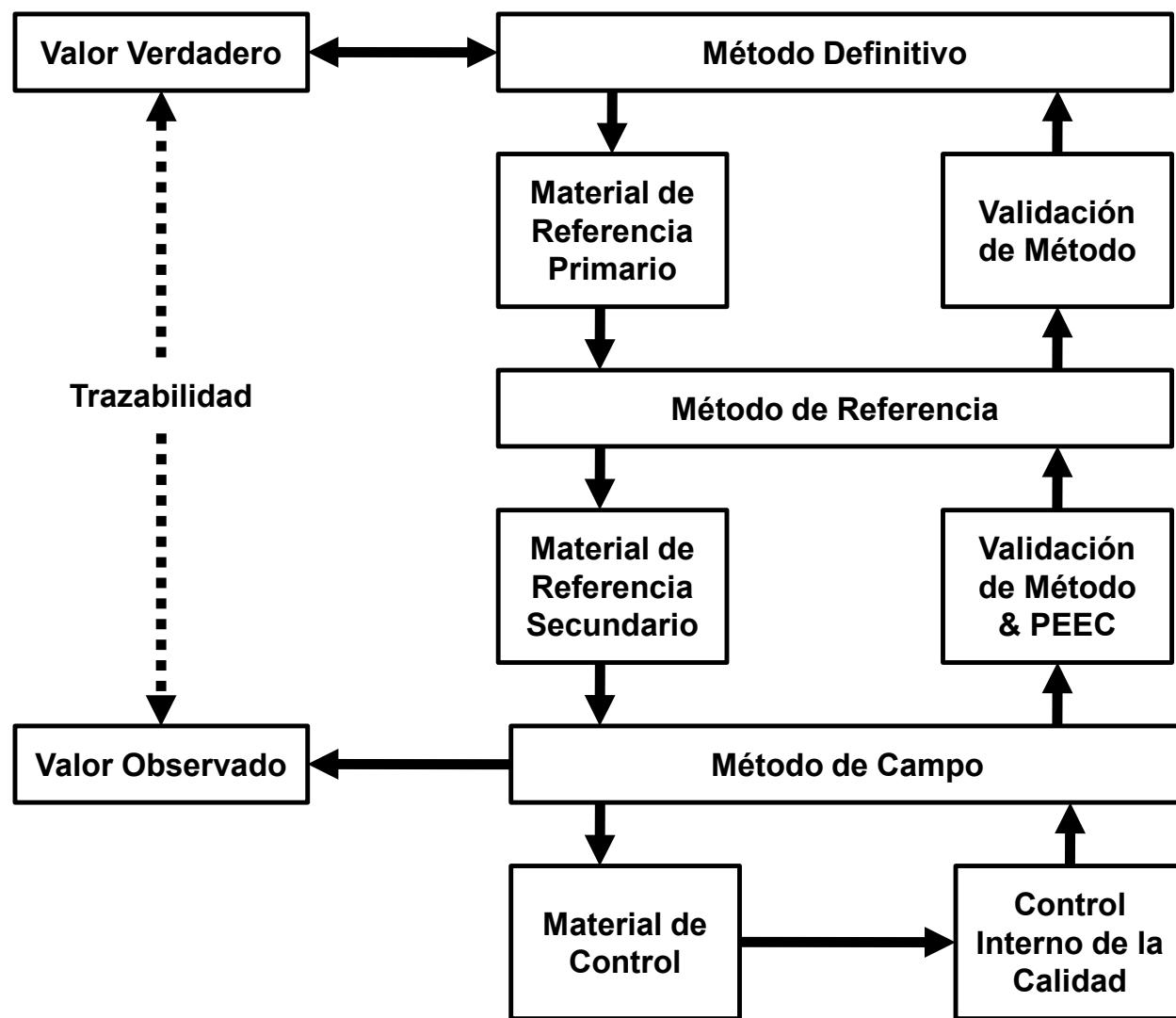
e incertidumbre de la medición. El resto de este capítulo esta dirigido a subsanar esta falencia dando una perspectiva de la dirección en que se dirigen los estándares globales y algunos cambios que es probable sean encontrados en un futuro cercano en los laboratorios de los E.U.A.

Veracidad y Trazabilidad

Primero vamos a ver algunas definiciones de la terminología ISO/GUM [1,2]:

- **Procedimiento de Examen:** conjunto de operaciones que tienen por objetivo determinar el valor o característica de una propiedad; (Nota: este es un término ISO que se corresponde con el de Método o procedimiento de medida);
- **Mensurando:** magnitud que se desea medir;
- **Exactitud de medida:** proximidad entre un valor medido y un valor verdadero de un mensurando;
- **Veracidad de medida:** proximidad entre la media de un número de valores infinitos medidos repetidos y un valor de referencia;
- **Precisión:** proximidad entre las indicaciones o los valores medidos obtenidos en mediciones repetidas de un mismo objeto, o de objetos similares, bajo condiciones especificadas;

Un principio fundamental es que la veracidad solo puede ser definida a través de la trazabilidad. La trazabilidad se refiera a un contexto de métodos de referencia y materiales de referencia que vinculan un método de prueba hacia atrás a la verdad o valor verdadero, como se muestra en la siguiente figura.



El punto de partida es un método definitivo que puede ser empleado para asignar valores a materiales de referencia primarios y que también genera la base de comparación para otros métodos de referencia más prácticos. Tales métodos de referencia y materiales de referencia secundarios se transforman en el vínculo práctico a la verdad para los métodos de campo, los cuales son implementados en los laboratorios de rutina y monitoreados a través del control estadístico interno de la calidad y control externo de la calidad. En este contexto, el control estadístico interno de la calidad se refiere al análisis de materiales de control por el laboratorio y a los esquemas de evaluación de la competencia o programas de evaluación externa de la calidad.

El colesterol es un ejemplo histórico sobre el desarrollo y la implementación de un sistema de referencia nacional en los E.U.A. El NIST (*US National Institute for Standards and Technology*) desarrolló un método definitivo

y un material certificado de referencia para colesterol. El CDC estableció una serie de laboratorios de referencia para soportar los esfuerzos nacionales de estandarización. Estos laboratorios empleaban el método de Abell-Kendahl para establecer una base de comparación para los métodos de campo. Los fabricantes brindan calibradores y métodos de campo que generan resultados consistentes en los servicios de laboratorio de rutina, dónde los métodos son monitoreados diariamente empleando material de control interno de la calidad estable y participan periódicamente en programas de evaluación externa de la calidad o esquemas de evaluación de la competencia.

Considerando estos sistemas de métodos y materiales de referencia, un valor trazable puede ser establecido por comparación como se muestra en la figura anterior.

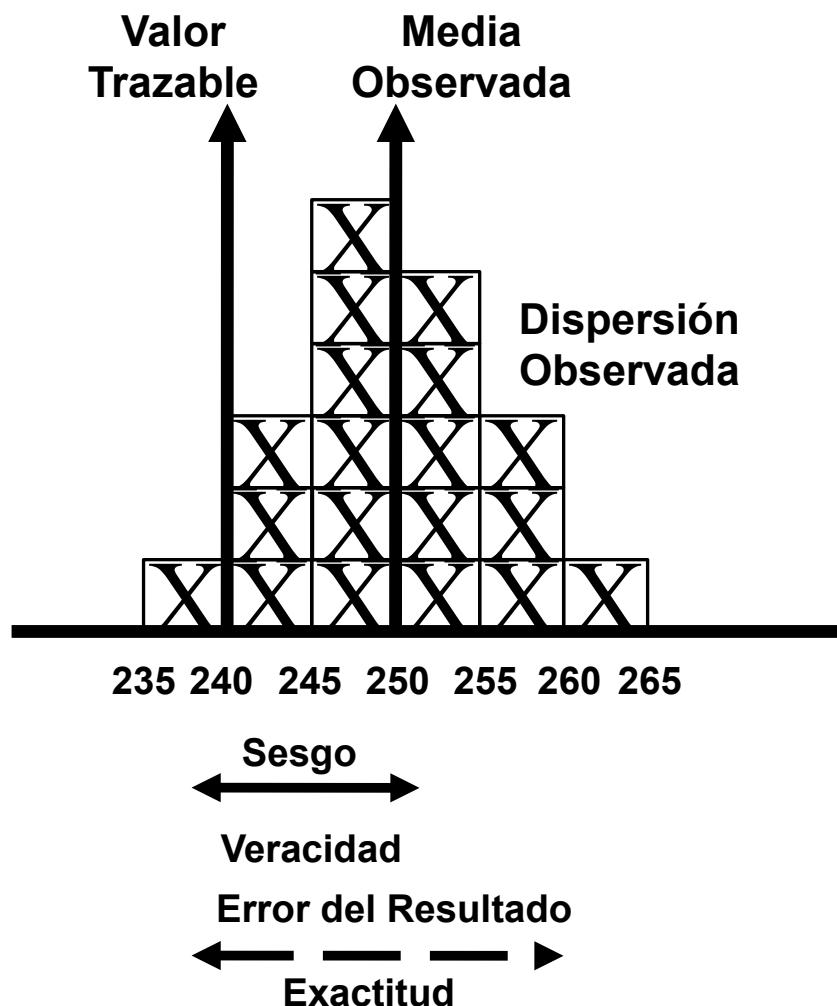
El concepto de valor trazable reemplaza al concepto anterior de valor verdadero que nunca pude ser conocido de manera exacta. Observe que ISO/GUM no define actualmente el término de valor trazable, pero asume que dicho valor puede sustituir al valor verdadero. Es evidente que el valor trazable no puede ser conocido exactamente y su significación debe ser descripta en términos de incertidumbre de medida (MU, por sus siglas en Inglés). En este nuevo mundo de términos de ISO y GUM, exactitud se transforma en el error de un resultado individual, lo cual tiene cierta similitud con el concepto de error total ya que puede ser afectado tanto por el error aleatorio como por el error sistemático, pero ahora son considerados (error aleatorio y sistemático) como fuentes de varianza, no como diferentes tipos de error.

Incertidumbre de Medida

Para comprender el concepto de incertidumbre de la medición es necesario familiarizarse con terminología adicional de ISO/GUM [1,2], como sigue:

- **Incertidumbre de medida:** parámetro asociado al resultado de una medición, que caracteriza la dispersión de los valores que podrían ser razonablemente atribuidos al mensurando;
- **Meta para la incertidumbre de medida:** Incertidumbre de medida formulada como meta u objetivo y decidida sobre la base de un uso previsto específico para la medición de un resultado;
- **Evaluación de Tipo A (de incertidumbre):** método de evaluación de la incertidumbre mediante análisis estadístico de series de observaciones;
- **Evaluación de Tipo B (de incertidumbre):** método de evaluación de la incertidumbre por medios distintos al análisis estadístico de series de observaciones;

ISO Veracidad y Exactitud



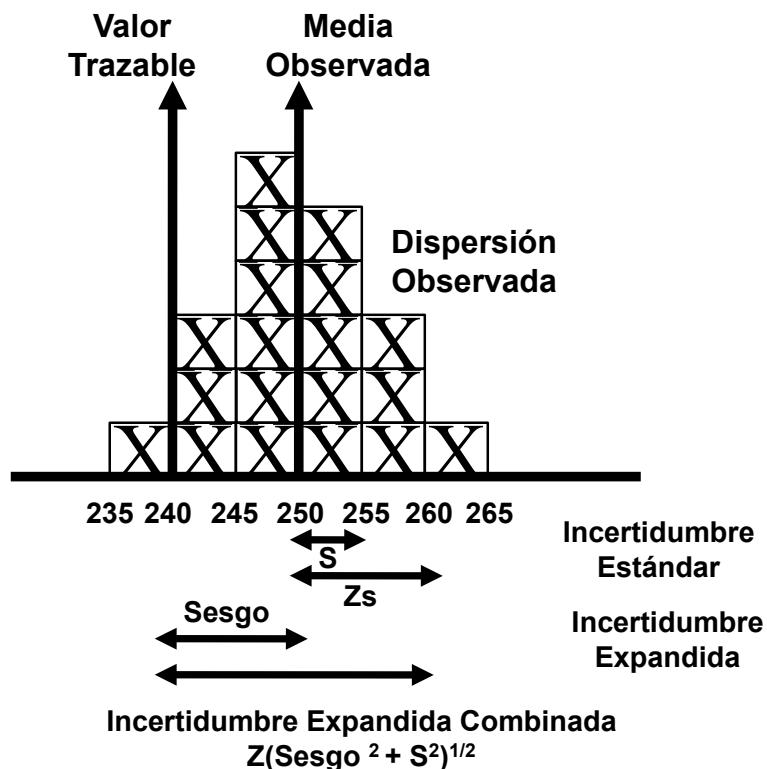
- **Incertidumbre Típica (Estándar):** incertidumbre del resultado de una medición, expresada en forma de desviación típica;
- **Incertidumbre Típica Combinada:** incertidumbre típica del resultado de una medición, cuando el resultado se obtiene a partir de los valores de otras magnitudes, igual a la raíz cuadrada positiva de una suma de términos, siendo éstos las varianzas o covarianzas de esas otras magnitudes, ponderadas en función de la variación del resultado de medida con la variación de dichas magnitudes;
- **Incertidumbre Expandida:** magnitud que define un intervalo vinculado al resultado de una medición, y en el que se espera encontrar una fracción importante de la distribución de valores que podrían ser atribuidos razonablemente al mensurando;

NOTA 1 La fracción puede entenderse como la probabilidad o el nivel de confianza del intervalo.

NOTA 2 Para asociar un nivel específico de confianza a un intervalo definido por la incertidumbre expandida, se requieren hipótesis explícitas o implícitas sobre la distribución de probabilidad representada por el resultado de medida y su incertidumbre típica combinada. El nivel de confianza que puede atribuirse a este intervalo posee la misma validez que las hipótesis realizadas.

NOTA 3 La incertidumbre expandida se denomina incertidumbre global en el apartado 5 de la Recomendación INC-1 (1980).

Los conceptos y terminologías vinculados a la incertidumbre están representados en la figura que se muestra a continuación. Una estimación de la incertidumbre puede darse en la forma de desviación estándar (incertidumbre Típica o Estándar). Estimados de componentes múltiples del proceso de medición pueden ser combinados adicionando las varianzas de los componentes individuales, luego tomando la raíz cuadrada de las varianzas combinadas (Incertidumbre típica combinada). Estos componentes pueden ser estimados experimentalmente (Evaluación del Tipo A de la incertidumbre) u obtenidos a partir de información publicada, teóricos, (Evaluación del Tipo B de la incertidumbre). Finalmente la incertidumbre puede ser expresada como un intervalo de confianza con un factor de cobertura establecido (una Incertidumbre Expandida o Incertidumbre Combinada Expandida con un factor de cobertura de 2 para



un intervalo del 95 %).

Los clientes y las aplicaciones previstas

Puede ser útil evaluar los conceptos de veracidad, exactitud, precisión, error total e incertidumbre de la medida en el contexto de los clientes quienes serán los que hagan uso de los conceptos. Diferentes clientes tendrán distintas capacidades y/o necesidades para aplicar los distintos conceptos. En el laboratorio clínico nosotros hemos empleado tradicionalmente el modelo de error para comprender la calidad de las pruebas y métodos. En nuestro mercado global cambiante, el modelo de incertidumbre puede no ofrecer alguna utilidad adicional para juzgar, controlar y manejar la calidad del proceso de examen del laboratorio. ¡Acá explicamos por qué!

En el mundo de la metrología, los clientes previstos son científicos de alto nivel quienes quieren conocer la confiabilidad del valor de un material certificado de referencia. En este entorno, los clientes previstos tienen el conocimiento del significado de la veracidad e incertidumbre de medida y están en condiciones de incorporar esa información dentro de su uso previsto.

¡Es completamente diferente en el mundo de la salud! Los médicos aún con un alto nivel de educación en las ciencias médicas, tienen poca comprensión sobre el significado de las mediciones analíticas y las características de desempeño de los procedimientos de medida. Ellos perciben que las pruebas de laboratorio brindan una información muy exacta, particularmente cuando los resultados están impresos con varias cifras significativas en informes generados por computadoras. Por lo tanto el laboratorio debe intervenir por el bien del médico así como también por el del paciente, brindando resultados que son confiables para el uso clínico previsto de la prueba.

De la misma manera los fabricantes deben intervenir en favor del laboratorio entregando procedimientos de medida con resultados comparables. Los laboratorios dependen de los fabricantes para brindar métodos con especificidad analítica, estandarización adecuada, y materiales de calibración que transfieran al laboratorio individual veracidad en sus resultados.

Como científicos del laboratorio clínico, nosotros debemos reconocer que es necesario en las pruebas de laboratorios en el ámbito de la salud manejar la calidad y alcanzar resultados comparables a partir de diferentes laboratorios. Es importante que esas actividades y responsabilidades estén adecuadamente enfocadas en las siguientes formas:

- La veracidad debe estar manejada primariamente por los fabricantes. Si se necesita alcanzar trazabilidad, será necesario un esfuerzo por parte

de los fabricantes. Pocos laboratorios tienen el tiempo y los recursos para hacer esto. En el laboratorio, la única aplicación práctica es para aquellos procedimientos de examen para los cuales existen materiales certificados de referencia y pueden ser utilizados para la trazabilidad de la calibración. La medida práctica es el sesgo lo cual indica que el concepto sistemático de exactitud está aplicado aquí.

- Exactitud, particularmente el sesgo residual del procedimiento de medida después de una estandarización apropiada, corrección del sesgo, y calibración apropiada en el campo, debe ser declarada por el fabricante y verificada por el laboratorio. La estimación práctica del sesgo hace uso del concepto sistemático de exactitud.
- La precisión debe ser declarada por el fabricante y verificada por el laboratorio. Como existen distintas condiciones experimentales para su estimación, el fabricante debe describir esas condiciones al momento de declarar sus especificaciones de precisión. El laboratorio debe considerar esas condiciones (declaradas por el fabricante) al momento de diseñar sus experimentos de verificación.
- El error total fue diseñado para evaluar rigurosamente los efectos del desempeño de los procedimientos de medida sobre la utilidad clínica de los resultados [4]. El error total debería ser manejado por los fabricantes y los laboratorios para establecer adecuación al uso previsto. Desafortunadamente las regulaciones de los E.U.A. no requieren que los fabricantes declaren sus especificaciones para la calidad, a través de la FDA parece que se estuviese alentando a las fabricantes a declarar una especificación de error total para la aprobación de los pruebas *waived*. No obstante los laboratorios deberían adoptar un criterio basado en el error total para tomar sus propias decisiones sobre la aceptabilidad de los procedimientos de medida de acuerdo a su aplicación prevista. Como se establece el documento de la CLSI EP 21 – A [5]:

“Se recomienda que para la mayoría de los casos, si uno tiene el conocimiento del error analítico total y resultados aberrantes, entonces uno tiene la información suficiente para evaluar la aceptabilidad de la prueba de diagnóstico.”

- La incertidumbre de medida aparentemente tiene algunas aplicaciones útiles para los científicos de laboratorios clínicos y otras para los médicos solicitantes. Los científicos del laboratorio clínico rara vez tienen las habilidades matemáticas de la gente que se dedica a la metrología, es por esto que la metodología recomendada por GUM no es práctica o útil para el laboratorio clínico. Los médicos rara vez tiene un entendimiento de la variación asociada a las mediciones, por lo tanto es probable que

aunque se les entregue una declaración sobre la incertidumbre de los resultados esta información no otorgue una ayuda real al momento de interpretarlos. En el mundo real es más probable que los médicos y los pacientes se sientan sorprendidos al ver que sus resultados tiene una incertidumbre asociada, desilusionados al saber que sus resultados tienen una duda asociada y preocupados al ver que sus resultados pueden ser erróneos. Es probable que los médicos no comprendan este concepto y que el mismo caiga en desgracia en el mercado. Por lo tanto la incertidumbre de medida debe ser dirigida por los clientes intermedios, primeramente los fabricantes y secundariamente el laboratorio.

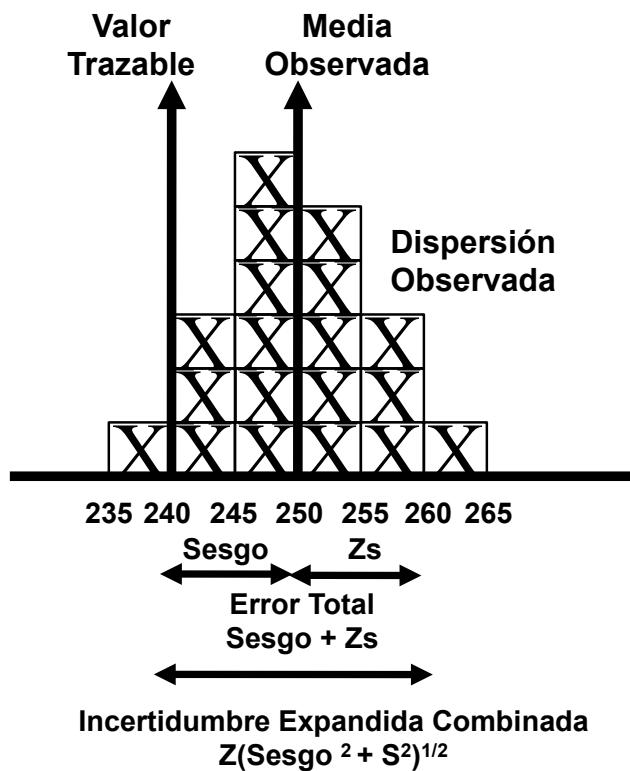
Por lo tanto, una estimación detallada de los componentes de la variación es importante para los fabricantes para mejorar la veracidad y confiabilidad de sus pruebas, precisión y exactitud son de utilidad para los fabricantes al momento de establecer y declarar sus especificaciones y para los laboratorios para verificar esas especificaciones de desempeño establecidas por los fabricantes. El error total es de utilidad para los científicos de laboratorio para juzgar sobre la aceptabilidad de sus métodos en relación a su uso clínico previsto y posiblemente a los médicos que están interesados en comprender la calidad promedio de las pruebas del laboratorio. ¡Los médicos comprenden el concepto de error!

Utilidad del Modelo de Error en el Mundo Real

Desde mi perspectiva, el error total fue el precursor de la incertidumbre de medida (actualmente se corresponde con el concepto de incertidumbre combinada expandida con un factor de cobertura de 2). Las similitudes se muestran en la figura más abajo, donde el concepto de incertidumbre se muestra como un modelo en la parte más baja del gráfico y puede ser comparado con el modelo de error total que se ve justo por arriba.

Aunque existen diferencias matemáticas en la forma en que los errores sistemáticos o sesgo son agregados a la varianza observada, o imprecisión, existe un propósito común que es el de mostrar el rango de valores que están asociados a un resultado individual. El modelo de estimación de la incertidumbre de medida de GUM es un modelo de máxima más detallado que caracteriza cada componente individual y luego los combina para determinar su efecto total. Este modelo es impracticable en los laboratorios clínicos y debería estar dirigido a los fabricantes y sus aplicaciones.

Guías existentes para valores diana o errores permitidos (requisitos de la calidad). Al momento de evaluar la confiabilidad de los procedimientos de medida tanto fabricantes como laboratorios tienen la responsabilidad de evaluar y verificar las características de desempeño del procedimiento de medida, identificar las fuentes de error o variación, y hacer una evaluación



sobre la aceptabilidad de esos errores o varianzas para el uso clínico previsto del procedimiento de medida. Acá es donde el modelo de error tiene una larga historia de aplicación en el laboratorio clínico y existen muchas guías y recursos para asistir a la definición de límites para error aceptable. Una conferencia de consenso internacional desarrollada en el año 1999 reconoció una jerarquía para las especificaciones de la calidad como se muestra a continuación [6]:

- Criterio de interpretación clínico, por ejemplo zona gris entre dos límites dónde se observan diferentes diagnósticos o distintas decisiones de tratamiento son tomadas,
- Metas biológicas para la imprecisión máxima permitida y el máximo sesgo permitido derivadas de la variabilidad biológica intra individuo y de grupo, las cuales también se combinan para establecer una meta para error total,
- Criterios o metas para error analítico total permitido para desempeño aceptable en un esquema de evaluación de la competencia (Proficiency Testing "PT") y Programas de evaluación externa de la calidad (PEEC),
- Opinión de grupos de expertos (por ejemplo, NCEP), y
- Guías sobre "estado del arte" basadas en el desempeño del método por ejemplo a partir de datos tomados de esquemas de evaluación de la competencia (Proficiency Testing "PT") y programas de evaluación externa de la calidad (PEEC).

Protocolos de Validación de métodos existentes. Ya existen prácticas estandarizadas, procedimientos experimentales y herramientas para el análisis de datos para evaluar el desempeño de un procedimiento de medida y validar su aceptabilidad considerando su uso previsto (contra requisitos de la calidad o valores diana para error permitido). La CLSI tiene disponibles una serie de documentos que han sido diseñados por comités técnicos, sujetos a consenso, revisión y aprobación y también sujetos a modificaciones posteriores y revisiones después de que son probados en el campo. Estos protocolos cubren las siguientes especificaciones de desempeño: precisión, veracidad o sesgo a partir de la comparación de métodos, interferencias, rango lineal, límites de detección e intervalos de referencia. Estos protocolos describen aproximaciones para el análisis de datos empleando herramientas estadísticas relativamente simples y por lo general incluyen hojas de trabajo que sirven de orientación para la recolección de datos y ejecución de cálculos haciendo de estos protocolos una herramienta práctica en el laboratorio hoy en día.

Comparabilidad con estrategias y métricas Seis Sigma. Una vez establecida la cantidad de error que puede ser permitida, existen herramientas y métricas que se ajustan a las estrategias de manejo de la calidad actual, particularmente la Gestión de la Calidad Seis Sigma. Los fabricantes han utilizado tradicionalmente índices sobre la “capacidad del proceso” para caracterizar el desempeño de un proceso relativo a ciertos “límites de tolerancia” o especificaciones de la calidad. En la década del 90 esos índices fueron transformados para describir a la calidad en una “escala sigma” y brindar así una herramienta de evaluación comparativa para la calidad entre las industrias y entre los procesos [7]. Las aplicaciones en el laboratorio han demostrado la utilidad de los conceptos y métricas Seis Sigma para asegurar la calidad de los procesos pre analíticos y pos analíticos [8], caracterización del desempeño y evaluación de la aceptabilidad de un nuevo procedimiento de medida [9], y la evaluación del desempeño entre laboratorios a partir de los datos obtenidos de esquemas de evaluación de la competencia (Proficiency Testing “PT”) y Programas de evaluación externa de la calidad (PEEC) [10].

Selección y diseño de procedimientos de Control Estadístico Interno de la Calidad. Un proceso crítico para la gestión de la calidad analítica es realizar el control estadístico interno de la calidad apropiado durante el servicio de operación de rutina del laboratorio. La norma ISO 15189 [1] identifica la necesidad de hacer esto en la sección 5.6.1 la cual enfatiza que el control estadístico interno de la calidad debe ser diseñado para “verificar la calidad prevista para los resultados de las pruebas”. Desafortunadamente el modelo de incertidumbre de medida no ha demostrado ofrecer alguna orientación sobre como cumplir con esta necesidad. Por otro lado, el modelo de error total, puede ser expandido para brindar modelos para la planificación de la calidad analítica

y clínica así también como herramientas para seleccionar las reglas de control correctas, y el número correcto de mediciones de control [11-12] sobre la base de la calidad requerida para la prueba y la precisión y exactitud observada para el procedimiento de medida. Por lo tanto el modelo de error total ya ha contribuido al desarrollo del proceso de planificación de la calidad y también ha aportado herramientas gráficas para una implementación práctica que ya son parte de la guía de consenso de la CLSI C24 A3 [13].

Considerando la variabilidad biológica intra individuo. Para muchas pruebas, la variabilidad biológica intra individuo puede ser muy grande, tan grande o aún más grande que la variabilidad analítica, por lo tanto la incertidumbre de la prueba debe tener en cuenta la variabilidad biológica propia del paciente. Por ejemplo, la variabilidad biológica intra individuo para el colesterol es aproximadamente del 6%, en comparación a la variabilidad analítica típica para el colesterol que es aproximadamente menor al 3%. Los modelos para establecer metas o requisitos de la calidad [14] o los modelos para la planificación de la calidad [12] ya han sido desarrollados y están disponibles. Existe y está disponible una base de datos con información sobre la variabilidad biológica intra individuo y de grupo, con metas de desempeño para precisión, sesgo y error total para unas 300 a 400 entidades diferentes [15].

Interpretación de los resultados de las pruebas en serie de pacientes. La variabilidad biológica intra individuo debe ser considerada al momento de interpretar una serie de resultados de un mismo paciente, es decir comparar un resultado frente a los antecedentes, lo cual es crítico para el seguimiento del tratamiento del paciente. Acá es donde el modelo de incertidumbre de medida falla en la dirección del uso previsto, no obstante exista ya una solución en el formato de “relación valores de cambio” (RCV, por sus siglas en Inglés), un término de incertidumbre definido por Fraser [14] para incluir la variabilidad analítica y la variabilidad biológica intra individuo de la siguiente forma:

$$RCV = 2^{1/2} * Z * (CV_A^2 + CV_I^2)^{1/2}$$

Donde

Z: Factor de Cobertura (1,96 para un 95 % o 2,33 para un 99 %)

CV_A: Imprecisión o Coeficiente de variación del procedimiento de medida

CV_I: Coeficiente de variación biológico intra individuo

Comunicación de los resultados a los médicos. Fraser [14] también ha desarrollado un modelo práctico para informar resultados y ayudar a los médicos a interpretar los cambios en los resultados de las pruebas frente a los límites de referencia, así como también los cambios frente a resultados anteriores

(antecedentes) usando los siguientes códigos en los informes:

- > Mayor que el límite de referencia
- < Menor que el límite de referencia
- >> Mayor que el límite de referencia y clínicamente significativo
- << Menor que el límite de referencia y clínicamente significativo
- * cambio significativos frente a antecedente (intervalo de confianza del 95%)
- ** cambio altamente significativo frente a antecedente (intervalo de confianza del 99%)

En efecto, se está reportando incertidumbre con este tipo de informes sin necesidad de generar números. En lugar de generar números, la estimación de la incertidumbre de medida se emplea para determinar si un resultado debe ser señalizado como con una desviación con respecto a los límites de referencia o si ha sufrido un cambio importante con respecto a un resultado inmediato anterior (antecedente). Esta aplicación toma en cuenta las necesidades del médico respecto a la información en un formato de fácil interpretación que puede ser fácilmente leído e interpretado por el cliente.

¿Cuál es el punto?

Veracidad, error total e incertidumbre de medida cumplen diferentes roles y a su vez son útiles a diferentes clientes. Los conceptos de veracidad e incertidumbre de medida deberían estar orientados a los fabricantes de dispositivos médicos y materiales de diagnóstico, quienes deben ser responsables de brindar procesos de examen y materiales trazables y confiables, y son responsables también de brindar pruebas cuyas resultados sean comparables método a método, laboratorio a laboratorio y de país en país.

El modelo de error, con el concepto de error total junto a las aproximaciones para establecer especificaciones para la calidad, es más útil a los científicos de laboratorio quienes deben evaluar el desempeño del método, implementar el control estadístico interno de la calidad, manejar la calidad del proceso de examen para asegurar que los resultados alcanzan la calidad para el uso clínico previsto de la prueba. Los conceptos y principios de ISO para sistemas de la calidad pueden ser aplicados en el laboratorio clínico como fue demostrado por Cooper y Gillions [16], pero el manejo de la calidad analítica debería seguir empleando el modelo de error junto con sus guías y herramientas que han sido desarrolladas específicamente para el laboratorio clínico.

Los científicos del laboratorio deben asumir la responsabilidad de la gestión de la calidad analítica para garantizar la calidad de los resultados que producen, no solo informar la incertidumbre y esperar que el médico solicitante y el paci-

ente hagan uso correcto de los resultados. Si se va a informar la incertidumbre, entonces el sistema para informar resultados debe ser cuidadosamente diseñado por los profesionales del laboratorio para ayudar a sus clientes, es decir, aplicar un sistema como el diseñado por Fraser [14] que señala los resultados que exceden los límites de referencia y alerta sobre cambios significativos en una serie de resultados (comparación frente antecedentes).

El argumento más fuerte en función de la necesidad de caracterizar la incertidumbre de medida será nuestra incapacidad para controlar las pruebas de laboratorio para “verificar la calidad prevista de los resultados de pruebas”, para usar la frase de la norma ISO 15189. ¡Si no conseguimos la calidad con eficacia, entonces tendremos que medir la calidad de estamos logrando, que puede ser caracterizada por la incertidumbre de medición!

Nuestro mundo está lleno de incertidumbre, para estar seguros, pero se vuelve más manejable en el laboratorio cuando la atención se centra en los errores analíticos.

Referencias

1. ISO/FDIS 15189 Medical laboratories – Particular requirements for quality and competence. 2002. International Organization for Standards, Geneva Switz. (Note that a 2nd edition of 15189 was released in 2007, and a third in 2012.)
2. GUM. Guide to the expression of uncertainty in measurement. ISO, Geneva, 1995.
3. CLSI. EP15-A2. User Verification of Performance for Precision and Trueness. Clinical Laboratory Standards Institute, Wayne, PA, 2005.
4. Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. Clin Chem 1974;20:825-833.
5. CLSI EP21-A. Estimation of total analytical error for clinical laboratory methods. Clinical Laboratory Standards Institute, Wayne, PA 2003.
6. Petersen PH, Fraser CG, Kallner A, Kenny D. Strategies to Set Global Analytical Quality Specifications in Laboratory Medicine. Scand J Clin Lab Invest 1999;59(7):475-585.
7. Harry M, Schroeder R. Six Sigma: The Breakthrough Management Strategy Revolutionizing the World's To Corporations. New York:Currency, 2000.
8. Nevalainen D, Berte L, Kraft C, Leigh E, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. Arch Pathol Lab Med 2000;124:516-519.
9. Westgard JO. Six Sigma Quality Design and Control: Desirable precision and requisite QC for laboratory measurement processes. 2nd ed. Madison WI:Westgard QC, 2006.

10. Westgard JO, Westgard SA. The quality of laboratory testing today: An assessment of sigma-metrics for analytical quality using performance data from proficiency testing surveys and the CLIA criteria for acceptable performance. *Am J Clin Pathol* 2006;125:343-354.
11. Westgard JO. Charts of operational process specifications (“OPSpecs charts”) for assessing the precision, accuracy, and quality control needed to satisfy proficiency testing criteria. *Clin Chem* 1992;38:1226-33.
12. Westgard JO. Internal quality control: planning and implementation strategies. *Ann Clin Biochem* 2003;40:593-611.
13. CLSI C24-A3. Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions; Approved Guideline – Third Edition. Clinical Laboratory Standards Institute, Wayne, PA, 2006.
14. Fraser C. Biological Variation: From Principles to Practice. AACC Press, 2001.
15. Ricos C, Alarez V, Cava F, et al. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999;59:491-500. [Note that the latest edition of this databank is available at <http://www.westgard.com/biodatabase1.htm>].
16. Cooper G, Gillions T. Producing Reliable Test Results in the Medical Laboratory: Using a quality system approach and ISO 15189 to assure the quality of laboratory examination procedures. Bio-Rad Laboratories, Quality Systems Division, Irvine CA 2007.

22: Respuestas de auto evaluación

Capítulo 1: ¿Es la calidad aún un asunto relevante para los exámenes de laboratorio?

¿Qué mitos acerca de la calidad existen en los laboratorios hoy en día?

Muchos analistas de laboratorio creen que la calidad, particularmente la calidad analítica, ya no es un tema relevante debido a que los laboratorios han implementado procedimientos de control de la calidad, planeamientos de garantía de la calidad y proyectos de mejora de la calidad. A menudo creemos erróneamente que:

- Los programas de garantía de la calidad aseguran la calidad en el cuidado de la salud.
- El control estadístico de la calidad controla la calidad de los exámenes de laboratorio.
- La calidad puede ser gestionada aun cuando se desconoce la calidad requerida.
- Los requisitos de la calidad necesitan considerar sólo la imprecisión y la inexactitud.
- Los métodos actuales tienen mejor imprecisión e inexactitud que la necesaria.
- Se considera hoy en día que la calidad analítica está asegurada.
- No es necesario aplicar grandes mejoras a nivel de la calidad analítica.
- El gobierno regula los exámenes de laboratorio para asegurar que la calidad es aceptable.
- Los laboratorios deben enfocarse en los errores pre y post analíticos porque los errores analíticos ya no son un problema.

¿Qué puede hacerse para mejorar las prácticas de manejo de la calidad en el laboratorio?

La calidad debe ser manejada de un modo objetivo y cuantitativo. Esto significa definir la calidad que es necesaria o deseada para cada examen, realizar estudios experimentales para validar que los métodos alcancen el desempeño necesario para producir la calidad deseada, y establecer los procedimientos de control de la calidad estadísticos para asegurar que la calidad deseada es alcanzada en la operación diaria.

Capítulo 2. ¿Cómo gestiona usted a la calidad?

¿Cuáles son los cinco componentes necesarios para gestionar a la calidad?

A menudo escuchamos acerca de las partes o componentes individuales de un sistema de la calidad, sin entender cómo encajan. El marco de la gestión total de la calidad (GTC) incluye procesos de la calidad de laboratorio (PCL), control de la calidad (CC), evaluación de la calidad (EC), mejora de la calidad (MC) y planificación de la

calidad (PC). Estos componentes deben funcionar como un ciclo o círculo de retroalimentación para proveer la mejora continua de la calidad.

¿Dónde encaja la validación de método en este marco GTC?

La validación de método (VM) debe ser un proceso estándar de laboratorio que se utiliza al implementar un nuevo método analítico. Como tal debería ser parte de los PCL – procesos de la calidad de laboratorio. Sin embargo, también debe entenderse que la VM es un proceso utilizado como parte de la planificación de la calidad, o PC, para proveer la evaluación técnica del desempeño del método que es necesaria para implementar nuevos métodos en el laboratorio.

¿Por qué es importante la validación de método?

Aunque los fabricantes prueben sus métodos extensivamente, hay muchos factores o variables que pueden cambiar cuando el método funciona en un laboratorio individual. Estos factores incluyen cosas como: diferentes números de lote de estándares y reactivos, cambios en los instrumentos o componentes del sistema, cambios en las condiciones ambientales en el laboratorio, diferentes niveles de habilidad de los operadores, y condiciones de transporte y almacenamiento que pueden afectar los reactivos y los materiales. La validación de métodos provee la garantía de que un nuevo método, sin importar los cambios que hayan ocurrido, aun opera aceptablemente bajo las condiciones de uso en el laboratorio. Es esencial para establecer los procesos de la calidad del laboratorio.

¿Qué guía o conduce el proceso de gestión de la calidad?

Los requisitos, metas y objetivos de la calidad son necesarios para conducir el proceso de gestión de la calidad y guiar la implementación y la evaluación de los servicios y procesos de examen del laboratorio.

Capítulo 3. ¿Cuál es el propósito de un estudio de validación de métodos?

¿Cuáles son los dos tipos principales de errores analíticos?

Los dos principales tipos de errores analíticos son la imprecisión y la inexactitud, los cuales pueden ser descritos como error aleatorio y error sistemático respectivamente. Los errores aleatorios pueden ser tanto positivos como negativos y su dirección no es predecible. Los errores sistemáticos son en una dirección y son predecibles.

¿Qué se entiende por “Error Total”?

El error total es el efecto neto de la imprecisión y la inexactitud en una medición individual o en un resultado de examen individual. Debido a que comúnmente se realiza una única medición cuando se hace un examen de rutina de laboratorio, esa medición puede ser errónea a causa de la imprecisión y la inexactitud del método. Realizar mediciones replicadas puede reducir el error aleatorio del resultado del examen, pero no reducirá el error sistemático o la inexactitud.

¿Cómo se relaciona el Error Total con los tipos de errores básicos?

El error total se estima comúnmente como el sesgo (o error sistemático) del método más un múltiplo (generalmente 2, 3 o 4) de la desviación estándar (o error aleatorio) del método. Por ejemplo, usando el criterio $ET = \text{sesgo} + 3S$, se espera que un resultado de un examen de colesterol será correcto dentro de 7.0 mg/dL si el error sistemático promedio o sesgo para un método de colesterol es 1.0 mg/dL y la S es 2.0 mg/dL. Esto indica que los estimados del sesgo y la S son apropiados para el nivel de concentración del espécimen de interés.

¿Cómo describe la literatura los errores de un método?

La mayoría de los estudios publicados abordan el tema de los estimados de precisión obtenidos a partir de estudios de replicación y los estimados de exactitud obtenidos partir de estudios de comparación de métodos. A menudo se documenta el rango reportable o el rango de desempeño lineal. Los límites de detección, la recuperación, interferencias y los intervalos de referencia son raramente incluidos. El informe quizás ofrezca un puñado de estadísticas sin ninguna intención de describir el tamaño de los errores.

¿Qué estadísticas se usan en un informe de validación de métodos?

Comúnmente encontrará promedios (o media), S y CV para resumir la información de un estudio de replicación. Para un experimento de comparación de métodos, verá regresión lineal o estadísticas de cuadrados mínimos (pendiente, intercepto-y y desviación estándar sobre la línea de regresión) y el coeficiente de correlación y a veces también estadísticas de exámenes “t” (sesgo, S de las diferencias y valores “t”). Estas son las diferentes estadísticas que deberá aprender para poder interpretar los resultados de los estudios de validación de método.

¿Cómo se vinculan las conclusiones del informe con los errores del método?

Las conclusiones acerca del desempeño de un método analítico son probablemente la parte menos estandarizada de un estudio de validación de método, aún en estudios publicados en prestigiosos artículos. Se debería relacionar el desempeño del método con los “estándares de desempeño” o “requisitos de la calidad” del examen.

Capítulo 4. ¿Cuáles son los requisitos regulatorios para la validación de métodos?

¿Cuáles son las clasificaciones de complejidad de un examen utilizadas para categorizar los requisitos de CLIA para la validación de métodos?

La Regla Final de CLIA clasifica los exámenes como microscopia realizada por el proveedor (PPM), *waived* y *non-waived*, donde la última clasificación incluye las categorías de complejidad moderada y alta de versiones anteriores de las regulaciones.

¿Cuáles son los requisitos de validación de métodos para los exámenes “waived”?

No hay requisitos de validación de métodos para estos exámenes. Los laboratorios solo necesitan seguir las instrucciones de los fabricantes.

¿Cuáles son los requisitos de validación de métodos para exámenes “non-waived” aprobados por la FDA?

Para exámenes “non-waived”, el laboratorio debe verificar las características de desempeño para rango reportable, precisión, exactitud e intervalos de referencia.

¿Cuáles son los requisitos de validación de método para los exámenes “non-waived” no aprobados por la FDA –o- modificados por el laboratorio?

Para los exámenes “non-waived” no aprobados por la FDA o modificados por el laboratorio, debe ser validada (además de rango reportable, precisión, y exactitud) la especificidad analítica (interferencias) y la sensibilidad analítica (detección de límite). Son requeridos estudios más extensivos, también, para establecer intervalos de referencia.

Capítulo 5. ¿Cómo es seleccionado un método?

¿Cuáles son los tres tipos de características de un método?

Las características de aplicación describen las condiciones que el método requiere si va a ser utilizado en el laboratorio. Las características de metodología son factores más técnicos relacionados al desempeño esperado del método. Las características de desempeño son mediciones específicas de desempeño que son importantes para que el método genere resultados aceptables.

Dé tres ejemplos de cada uno de los tipos de características de un método.

Los ejemplos de características de aplicación pueden ser tipo de espécimen, tamaño de la muestra y costo. Si un método no satisface las condiciones necesarias, su uso no debe ser considerado en el laboratorio. Los ejemplos de características de metodología pueden ser el tipo de estándar utilizado (primario contra secundario), el principio de medición (fotométrico contra electro-químico), o la reacción química específica elegida (enzimática contra no enzimática). Los ejemplos de características de desempeño son el rango reportable, la precisión, la exactitud, la recuperación la interferencia y el límite de detección.

¿Qué características son de mayor interés al momento de seleccionar un método?

Las características de aplicación deben ser cumplidas en la selección inicial de un método. Si el método no trabaja con el tipo correcto de espécimen, no tiene un tamaño de muestra lo suficientemente pequeño, o el costo es excesivo, entonces no hay razón para probar su desempeño. Las características de metodología son, del mismo modo, útiles para guiar la selección hacia un método del que se espera el desempeño deseado.

¿Cuáles son las características de mayor interés durante la validación de un método?

Los estudios de validación de método son los que realmente evalúan las características de desempeño. La precisión y la exactitud son las características de desempeño fundamentales de cualquier método y deben siempre ser evaluadas. Otras características de desempeño importantes incluyen el rango reportable, la recuperación, la interferencia y el límite detección.

Capítulo 6. ¿Qué experimentos son necesarios para validar el desempeño de un método?

¿Qué experimento es utilizado para estimar la imprecisión de un método?

Un experimento de replicación se realiza haciendo diversas mediciones, generalmente un mínimo de 20, en una muestra que es estable por un período de tiempo de interés en el estudio. Típicamente, al comienzo se estudia por un período corto de tiempo, 1 corrida o 1 día, luego se estudia un período más largo, 20 corridas a lo largo de días.

¿Qué experimentos son utilizados para estimar la inexactitud de un método?

El experimento de comparación de métodos es comúnmente el más utilizado para proveer un estimado total del “sesgo” o del error sistemático entre el nuevo método, o método de evaluación, y un método de comparación establecido. Los experimentos de recuperación e interferencia también pueden ser utilizados para evaluar si un material específico o la matriz de una muestra causan errores sistemáticos.

¿Por qué el experimento para rango reportable es realizado al comienzo del estudio de validación?

Este experimento define el rango analítico que se va a considerar para juntar datos e información para los demás experimentos. Por ejemplo, las concentraciones de los materiales seleccionados para el experimento de replicación deberían estar dentro del rango (reportable) útil del método. Los especímenes de pacientes en el estudio de comparación de métodos deberían ser seleccionados para estar en el rango reportable del método.

¿Por qué por lo general se realizan dos experimentos de replicación diferentes?

La información más rápida acerca de la precisión se obtiene de un estudio a corto plazo, como el de intra corrida o intra día. Si esta estimación muestra que el desempeño del método no es aceptable, no hay razón para continuar y realizar otros estudios. Si el desempeño a corto plazo es aceptable, entonces se necesitará un estudio a largo plazo para evaluar el desempeño bajo condiciones que representarán más realísticamente lo que es esperado durante el servicio de rutina del laboratorio.

¿Qué diferencia existe entre el error sistemático proporcional y constante?

Un error constante es aquel cuyo tamaño en unidades de concentración permanece igual aun cuando el valor real del resultado del examen cambia. Un error proporcional es aquel cuyo tamaño en unidades de porcentaje permanece igual cuando el valor real del resultado del examen cambia. Por ejemplo, al calibrar un método, un error al ajustar el cero producirá generalmente un error constante, mientras que un error al ajustar la calibración producirá un error proporcional.

¿Qué experimentos se realizan para estimar los errores sistemáticos proporcional y constante?

Los experimentos de interferencia pueden ser usados para estimar los errores sistemáticos constantes (EC) a causa de los materiales específicos. Los estudios de recuperación pueden ser usados para estimar los errores sistemáticos proporcionales (EP). También puede obtenerse información acerca de los errores sistemáticos constantes y proporcionales a través de un experimento de comparación de métodos al interpretar cuidadosamente las estadísticas de regresión (la pendiente es relacionada al EP, el intercepto-Y al EC).

Capítulo 7. ¿Cómo se analizan los datos experimentales?

¿Qué cálculos realizaría usted con los datos obtenidos a partir del experimento de replicación?

Las estadísticas generalmente calculadas son la media, la S y CV. Ocasionalmente, los valores “F” son calculados para comparar las S observadas de los diferentes métodos.

¿Qué cálculos realizaría usted con los datos obtenidos a partir del experimento de comparación de métodos?

Generalmente se calculan las estadísticas de regresión lineal (pendiente, intercepto-Y, desviación estándar de los puntos sobre la línea de regresión). A menudo se incluye el coeficiente de correlación. Otros cálculos que también suelen encontrarse son las estadísticas para datos apareados de exámenes “t” (sesgo, desviación estándar de las diferencias, valor “t”).

¿Qué gráficos están asociados a los diferentes tipos de análisis?

Un histograma provee un útil gráfico de la información de un estudio de replicación. Para la información proveniente de un experimento de comparación de métodos, comúnmente se usa un gráfico de comparación (valor del examen como coordenada “y”, valor comparativo como coordenada “x”) cuando se calculan las estadísticas de regresión. Cuando las estadísticas “t” son calculadas a menudo se utiliza un gráfico de diferencias (la diferencia entre valor obtenido por el método de examen y el valor obtenido por el método de comparación como la coordenada en “y”, contra el valor obtenido por el método de comparación como la coordenada en “x”).

Estime la media, desvío estándar y coeficiente de variación a partir del siguiente conjunto de datos:

[203, 202, 204, 201, 197, 200, 198, 196, 206, 198, 196, 192, 205, 195, 207, 198, 201, 195, 202, 195]

La razón para hacer esta pregunta (y la siguiente) es saber si tiene acceso a herramientas de cálculo tales como una calculadora de mano, una planilla electrónica de cálculo o un programa estadístico para trabajar en un ordenador.. También puede hacer uso de las calculadoras *online* disponibles en:

<http://www.westgard.com/mvtools.html>

Para este conjunto de datos (que también es el conjunto de datos ejemplo provisto con las calculadoras en línea nombradas más arriba) la media es 199,55, la S es 4,1861 y CV es 2,0977. En la práctica, redondearíamos la S a 4,19 y CV a 2,10%.

Estime los estadísticos de regresión y el sesgo promedio a partir del siguiente conjunto de datos:

Método de Examen (Y)	Método de Comparación (X)
218	217
161	161
240	244
193	193
290	295
117	118
118	122
203	204
74	74
114	116
245	238
262	260
203	203
218	207
311	304
362	353
332	327
428	423
163	155
268	257

Este conjunto de datos es también la información disponible a manera de ejemplo en la calculadora de datos apareados *online* disponible en la dirección mencionada arriba. La pendiente es 1,0226, el intercepto-Y es -2,6022, la S de los puntos sobre la línea de regresión es 4,754 y el coeficiente de correlación es 0,9986. En la práctica, redondearíamos las estadísticas de regresión a 1,023, -2,60 y 4,75 respectivamente.

Capítulo 8. ¿Cómo se calculan estas estadísticas?

En un estudio de validación de métodos de colesterol, el estadístico-t obtenido a partir de la comparación de métodos arroja los siguientes datos:

Sesgo = 1.74 mg/dL, Sdif = 5.90 mg/dL, N = 81, y t = 2.65. ¿Cuál es el valor crítico de “t” para p = 0.05? ¿Es el sesgo observado estadísticamente significativo? ¿Es el sesgo observado clínicamente significativo?

Para N=81, el valor “t” crítico puede ser encontrado en el cuadro y es aproximadamente 1,99 para P=0,05 (examen bilateral de significancia, df=N-2). Nótese que a menudo tendrá que interpolar entre los valores dados en los cuadros de valores críticos. El valor “t” observado de 2,65 es mayor que el valor crítico de 1,99, por lo tanto el sesgo observado es estadísticamente significativo o “real”. Sin embargo, la magnitud del sesgo (1,74 mg/dL) podría ser lo suficientemente pequeña para ser aceptable para el uso clínico de un método. Es por eso que no recomendamos usar un examen “t” para evaluar la aceptabilidad del desempeño del método. En vez de eso recomendamos una técnica gráfica conocida como el Cuadro de Decisión de Método, descrito en el capítulo 16.

En un estudio de validación de métodos, el método en evaluación tiene una desviación estándar de 4.0 mg/dL y el método de comparación tiene una desviación estándar de 5.0 mg/dL, ambos valores fueron estimados a partir de 21 mediciones efectuadas sobre el mismo material de control cuya concentración aproximada es de 200 mg/dL. ¿Cuál es el valor calculado para el estadístico “F”? ¿Es la diferencia de desempeño a nivel de precisión estadísticamente significativa? ¿Es la diferencia de desempeño a nivel de precisión clínicamente significativa?

El valor “F” calculado es de 1,56 (25/16). Para 20 grados de libertad (df=N-1) para cada método, el valor crítico “F” puede encontrarse en el cuadro y es de 2,12. Debido a que el valor “F” observado es menor que el valor “F” crítico, la diferencia no es estadísticamente significativa. Sin embargo, la significancia clínica de la imprecisión observada depende del tamaño de la S del método comparada con la cantidad de error permitido. Nuevamente, no recomendamos el uso de un examen de significancia, en este caso el examen “F”, para juzgar la aceptabilidad del desempeño de un método. El Cuadro de Decisión de Método le permitirá considerar la precisión y la exactitud del método en relación a un requisito de la calidad definido en la forma de la cantidad de error total que es permitido para el examen.

Para verificar las especificaciones de precisión declaradas por el fabricante se calculó el estadístico “F” para comparar la desviación estándar de 5.0 mg/dL obtenida en un experimento de replicación para el método evaluado a partir de un N= 31 con la desviación estándar declarada por el fabricante de 4.0 mg/dL obtenida también de un experimento de replicación con N= 31. ¿Ha verificado usted las especificaciones del fabricante para precisión?

El valor “F” calculado es nuevamente 1,56 (25/16). Para 30 grados de libertad (N-1) por ambos métodos, el valor crítico “F” es de 1,93. Debido a que el valor “F” cal-

culado es menor que el valor “F” crítico, la información no muestra que la diferencia sea estadísticamente significativa. En este caso la precisión observada en el laboratorio no es diferente a la declarada por el fabricante, de esta manera la especificación del fabricante ha sido verificada.

Para verificar las especificaciones de precisión declaradas por el fabricante se calculó el estadístico “F” para comparar la desviación estándar de 5.0 mg/dL obtenida en un experimento de replicación para el método evaluado a partir de un N= 11 con la desviación estándar declarada por el fabricante de 3.0 mg/dL obtenida también de un experimento de replicación con N= 11. ¿Ha verificado usted las especificaciones del fabricante para precisión?

El valor “F” calculado es 2,78 (25/9). El valor crítico “F” para 10 grados de libertad (N-1) por ambos métodos es de 2,98. Debido a que el valor “F” calculado es menor que el valor “F” crítico, no hay diferencia estadísticamente significativa entre la precisión observada en el laboratorio y aquella declarada por el fabricante. Por lo tanto la información no es inconsistente con la especificación declarada por el fabricante y puede decirse que la especificación del fabricante ha sido verificada. Nótese, sin embargo, que si se hubiese realizado un experimento de mayor peso con 31 mediciones, el valor crítico “F” sería 2,70 y la información obtenida nos diría que la especificación del fabricante no fue verificada. Este es un ejemplo que muestra que tener poca información limita la capacidad de evaluar adecuadamente el desempeño de un método frente a la especificación declarada por el fabricante.

Para verificar las especificaciones del fabricante para exactitud se trabajó con el estadístico “t”, se obtuvo un sesgo de 1.5 mg/dL y un valor de “t” de 2.5 para un estudio donde se procesaron 40 muestras de pacientes por el método a evaluar y el mismo método de comparación empleado por el fabricante al momento de establecer sus especificaciones. El fabricante declaró que no existía sesgo entre ambos métodos. ¿Ha verificado usted las especificaciones declaradas por el fabricante para exactitud?

El valor crítico “t” es 2,02 para un N de 40. Debido a que el valor t calculado de 2,50 es mayor que el valor crítico “t” de 2,02, la diferencia es estadísticamente significativa o “real”. La información no verifica la especificación declarada por el fabricante que dice que el sesgo debería ser cero entre estos dos métodos. Sin embargo, el tamaño del sesgo es pequeño (1,5 mg/dL) y el desempeño del método podría ser aceptable para los propósitos clínicos. Una vez más, este ejemplo muestra que las conclusiones acerca de la significancia estadística y la significancia clínica pueden ser diferentes. Nuevamente, una mejor propuesta es usar el Cuadro de Decisión de Método en vez de depender de exámenes de significancia estadística.

Capítulo 9. ¿Cómo se determina el rango reportable de un método?

¿Qué cantidad de niveles de diferente concentración son necesarios para validar el rango reportable?

El mínimo es 4, pero normalmente se prefiere 5 o 6.

¿Cuál es una forma práctica de obtener una serie de esos materiales?

Una serie de 5 materiales puede ser fácilmente preparada usando un pool de pacientes “alto” y uno “bajo”. Mezcle 3 partes del pool “bajo” con 1 parte del pool “alto”, 2 partes del “bajo” con 2 partes del “alto” y 1 parte del “bajo” con 3 partes del “alto”. Estas 3 mezclas junto con los pools originales “alto” y “bajo” le dará una serie de 5 muestras.

¿Cuántas repeticiones se efectúan por lo general sobre cada uno de los materiales?

Un mínimo de 2, pero es preferible que se realicen 3 o 4.

¿Cómo se analizan los datos para evaluar el rango reportable?

A pesar de todas las estadísticas que pueden ser calculadas, el estándar de oro para evaluar el rango reportable es graficar los valores observados o el promedio de los replicados en el eje “y” contra el valor asignado para los materiales en el eje “x”. Dibuje la mejor línea recta a través de la información, asegurándose que la línea pase a través de los puntos al final inferior del rango. Inspeccione visualmente el gráfico y determine el rango lineal donde los resultados de los exámenes serán confiables. Describa este “rango reportable” por el límite inferior (generalmente cero) y el límite superior determinado por el gráfico obtenido a partir de sus datos.

¿Cómo se analizan los datos para evaluar linealidad?

La linealidad no es realmente la característica de desempeño que necesita ser validada! Lo que es de interés es el rango reportable. Sin embargo, puede usar estadísticas de regresión lineal o de primer orden para ajustar la información y describir estadísticamente la línea observada y el ajuste de la información. Existen exámenes estadísticos más elaborados y más complicados publicados en la literatura para evaluar la linealidad, pero por razones prácticas, la mayoría de la gente solo inspecciona visualmente la información y la línea de mejor ajuste para determinar la linealidad.

Capítulo 10. ¿Cómo se determina la imprecisión de un método?

¿Cuántos niveles o materiales se necesitan para un estudio de replicación?

Un mínimo de dos diferentes niveles o concentraciones, y preferiblemente tres para brindar estimados a niveles bajos, normales y elevados.

¿Cuál es el número mínimo de mediciones que generalmente se recogen?

Comúnmente se acepta que el número mínimo de mediciones replicadas debería ser de 20. Sería mejor realizar más mediciones debido a la dificultad para obtener un estimado confiable de la S aun con 20 mediciones.

¿Qué período de tiempo debería abarcar el experimento de replicación final?

La mayoría de los estudios adhieren a un período de tiempo mínimo de 20 días

hábiles o un mes de calendario debido al alto costo de extender el período de tiempo del estudio más allá del mínimo requerimiento.

¿Cómo se analizan los datos para estimar la imprecisión?

Calcule la media, la S y los CV para cada uno de los materiales. Prepare un histograma para documentar la forma de la distribución.

Capítulo 11. ¿Cómo se determina la inexactitud o sesgo de un método determinado?

¿Cuál es el número mínimo de especímenes que se utilizan generalmente en la comparación?

Se recomienda generalmente un número de 40. Sin embargo, se puede obtener un resultado aceptable con un mínimo de 20 si los especímenes han sido seleccionados para cubrir de manera homogénea del rango analítico del procedimiento de medida evaluado. Si los especímenes se eligen de manera aleatoria, muchos valores de los exámenes caerán en el intervalo de referencia para la población saludable y por lo tanto se necesitarán más especímenes para cubrir el rango analítico amplio deseado para el experimento.

¿Cuántos replicados de las mediciones se efectúan por lo general?

Se prefieren efectuar duplicados por el método de examen y el método de comparación. Sin embargo, es común realizar mediciones únicas para ambos métodos cuando estos han sido bien estudiados por los fabricantes o bien documentados por los estudios publicados.

¿Qué período de tiempo se recomienda para la duración del experimento?

Se recomienda un mínimo de 5 días para los experimentos de comparación de métodos con el fin obtener datos representativos del desempeño de los métodos. Un período de tiempo mayor es aceptable. Es frecuente realizar el estudio de comparación de métodos en paralelo con el estudio de replicación a largo plazo, el cual usualmente cubre un período de 20 días.

¿Cómo deberían ser calculados los datos cuando existen dos niveles de decisión médica?

Es conveniente utilizar estadísticas de regresión lineal cuando hay dos o más niveles de decisión de interés. Estas estadísticas permiten estimar el error sistemático en cualquier concentración dentro del rango analítico que ha sido estudiado.

¿Cómo deben graficarse los datos si hay dos o más niveles de decisión médica?

Usando un gráfico de comparación donde el resultado obtenido por el método de examen es graficado en el eje "y" contra el resultado obtenido por el método comparativo en el eje "x".

¿Qué cálculos deben ser utilizados si hay un único nivel de decisión médica?

Las estadísticas de regresión lineal pueden usarse para calcular el error sistemático en un nivel de decisión, pero también es posible que las estadísticas de los exámenes “t” brinden un estimado confiable del error sistemático si la media de la información es cercana al nivel de decisión de interés. Si el rango de la información analítica es estrecho, entonces se preferirán las estadísticas de exámenes “t” y no las estadísticas de regresión.

¿Cómo se grafican los datos cuando existe un único nivel de decisión médica?

Junto con el uso de estadísticas de exámenes “t”, puede usar un “gráfico de diferencias” en el cual la diferencia entre el resultado obtenido por el método de examen y el resultado obtenido por el método de comparación se grafica en el eje “y” contra el resultado obtenido a partir del método de comparación que se representa en el eje “x”.

Capítulo 12. ¿Cómo usar la estadística para estimar errores analíticos?

El siguiente resumen estadístico fue obtenido a partir de un experimento de comparación de métodos para glucosa:

$$a = 5,23 \text{ mg/dL}, b = 0,999, Sy/x = 7,23 \text{ mg/dL}, \text{sesgo} = 5,13 \text{ mg/dL}, S_{\text{dif}} = 7,23 \text{ mg/dL}, t = 8,03, r = 0,996, N = 128.$$

¿Cuál es el error sistemático proporcional entre los métodos?

Hay muy poco error proporcional dado que la pendiente es de 0,999, lo cual es solo 0,001 o 0,1% diferente al valor ideal de 1,000.

¿Cuál es el error sistemático constante entre los métodos?

El error constante puede ser estimado del intercepto-Y o del sesgo, que son de 5,23 mg/dL y 5,13 mg/dL respectivamente.

¿Cuál es el error aleatorio entre métodos?

El error aleatorio entre los métodos puede ser estimado a partir de Sy/x o del S_{dif} cuyos valores son de 7,23 mg/dL.

¿Por qué hay tan buena concordancia entre las estimaciones del error obtenido a partir de estadísticas de regresión y estadísticos “t”?

Los estimados de errores por regresión y los exámenes “t” tienen tan buena concordancia porque no hay error proporcional presente.

¿Es el error sistemático entre métodos estadísticamente significativo o “real”?

Sí, el error sistemático es estadísticamente significativo o real, como lo muestra el hecho de que el valor “t” calculado de 8,03 es mayor que el valor “t” crítico, que es aproximadamente 2,0 cuando N es 30 o más.

¿Qué le dice el coeficiente de correlación?

El alto valor para el coeficiente de correlación indica que se ha obtenido una amplia gama de información y los estimados de regresión de la pendiente y del intercepto deberían ser confiables.

El siguiente resumen estadístico fue obtenido a partir de un experimento de comparación de métodos para nitrógeno ureico:

$N = 316$, $a = -0,31 \text{ mg/dL}$, $Sa = 0,23 \text{ mg/dL}$, $b = 1,032$, $Sb = 0,009$, $Sy/x = 0,97 \text{ mg/dL}$, $Sx = 13,2 \text{ mg/dL}$, $r = 0,997$, $\text{sesgo} = 0,40 \text{ mg/dL}$, $S_{\text{dif}} = 1,08 \text{ mg/dL}$, $t = 6,58$.

¿Cuál es el error sistemático proporcional entre los métodos?

Hay un error sistemático proporcional de alrededor de 3,2% como se revela por la diferencia entre la pendiente observada de 1,032 y el valor ideal de 1,00.

¿Cuál es el error sistemático constante entre los métodos?

El mejor estimado del error constante está dado por el intercepto-Y (-0,31 mg/dL); en este caso hay un error proporcional entre los métodos, lo que hace sospechar las estadísticas de la prueba t.

¿Por qué es mejor usar las estadísticas de regresión para estimar los errores en lugar de usar un estadístico t?

Recuerde que los estimados de errores a través de las estadísticas de pruebas t son influenciadas por el error proporcional, que se evidencia tanto por el valor del Sesgo como por el valor de SDdiff, aun cuando no es un error de este tipo. Por lo tanto es mejor estimar los componentes constantes y proporcionales del error sistemático usando la pendiente y el intercepto y respectivamente.

¿Cuál es el intervalo de confianza del 95% para la intersección-Y?

¿Difiere la intersección-Y significativamente del valor ideal de 0,0?

El intervalo de confianza del 95% para el intercepto-Y puede ser calculado a partir del valor del intercepto-Y (a) \pm 2 veces la desviación estándar del intercepto-Y (Sa), en el cual “ a ” es -0,31 mg/dL, “ Sa ” es 0,23 mg/dL y 2^*Sa es 0,46 mg/dL. Esto significa que el intervalo de confianza del 95% para el intercepto-Y va desde -0,76 mg/dL hasta +0,15 mg/dL. Debido a que este intervalo de confianza incluye el valor ideal de 0,0, el intercepto-Y no es significativamente diferente del valor ideal.

¿Cuál es el intervalo de confianza del 95% para la pendiente?

¿Difiere la pendiente significativamente del valor ideal de 1,0?

El intervalo de confianza del 95% para la pendiente puede ser calculado del valor de la pendiente (b) \pm 2 veces la desviación estándar de la pendiente (Sb), en el cual “ b ” es 1,032, “ Sb ” es 0,009 y 2^*Sb es 0,018. Esto significa que el intervalo de confianza del 95% para la pendiente va desde 1,014 hasta 1,050. Debido a que este rango no

incluye el valor ideal para la pendiente de 1,0, puede decirse que la diferencia entre la pendiente observada e ideal es estadísticamente significativa.

¿Qué le dice el coeficiente de correlación?

El alto valor del coeficiente de correlación indica que se ha obtenido una amplia gama de información y que los estimados de la pendiente y el intercepto deberían ser confiables.

Capítulo 13. ¿Cómo examinar fuentes específicas de inexactitud?

¿Cuál es la diferencia entre un experimento de recuperación y uno de interferencia?

Los experimentos se ven similares, pero son en realidad diferentes. En un experimento de recuperación, se añade el analito que el examen está midiendo. En un experimento de interferencia, se añade un material diferente que potencialmente interferirá en la medición del analito de interés. Los cálculos son bastante diferentes porque los experimentos se utilizan para estimar diferentes tipos de errores analíticos. El experimento de recuperación se usa para estimar los errores sistemáticos proporcionales, por lo tanto los resultados son en porcentaje. El experimento de interferencia se usa para estimar errores sistemáticos constantes, por lo tanto los resultados son en unidades de concentración.

¿Qué interferencias comunes se estudian usualmente?

Los efectos provocados por elevaciones en los niveles de bilirrubina y la presencia de lipemia y hemólisis, son comúnmente estudiados por la mayoría de los métodos.

¿Cuál es la manera apropiada de calcular la recuperación?

Debido a que el error proporcional debería ser representado como un porcentaje, los resultados de un experimento de recuperación son proporcionados para brindar una proporción que luego puede ser convertida a un porcentaje. La cantidad recuperada debe ser dividida por la cantidad añadida. La cantidad recuperada es la diferencia entre el resultado obtenido para la muestra añadida y el resultado obtenido para la muestra empleada como blanco. Un error común es dividir el resultado de la muestra añadida por el resultado de la muestra de blanco, en vez de calcular la diferencia y dividir por la cantidad añadida.

Capítulo 14. ¿Cuál es el valor más bajo confiable para un método?

¿Cuándo usted necesita realizar un experimento de límite de detección?

El límite de detección debería ser validado para aquellos exámenes en los cuales los valores inferiores son interpretados críticamente. El límite de detección es importante para las drogas forenses con el fin de documentar la presencia o ausencia de las drogas. El límite de detección también es importante para marcadores tumorales que se usan para dar seguimiento a los resultados de un tratamiento.

¿Qué estimación del límite de detección dará generalmente el valor más alto?

El Límite de Cuantificación “LC” (LoQ, por sus siglas en inglés) dará generalmente el estimado más alto del límite de detección. La Sensibilidad Funcional (FS, por sus siglas en inglés) dará el estimado próximo más alto. El Límite de Blanco (LoB, por sus siglas en inglés) dará el más bajo.

¿Por qué el Límite de Detección “LB” (LoD, por sus siglas en inglés) es siempre más alto que el Límite Vacío (LoB)?

El Límite de Detección (LoD) es siempre más alto que el Límite de Blanco (LoB) porque es definido como LoB +1,65 • Sspk, en el cual Sspk es la desviación estándar observada para una muestra suplementada de bajo nivel.

Capítulo 15. ¿Cómo se verifica el intervalo de referencia de un método?

¿Cuántos especímenes de pacientes se necesitan generalmente para verificar un intervalo de referencia?

De acuerdo a los lineamientos de CLSI, se necesitan un mínimo de 20 especímenes para verificar la transferencia de un intervalo de referencia recomendado por el fabricante al laboratorio.

¿Cuál es el número mínimo de especímenes de pacientes para establecer un intervalo de referencia?

Los lineamientos de CLSI por lo general recomiendan obtener 120 individuos que representen un grupo de interés. Se necesita un mínimo de 60 para brindar un estimado preliminar.

¿Cómo se puede transferir un intervalo de referencia desde un método de comparación?

Asumiendo que en el pasado ha habido estudios adecuados para estimar el/los intervalo/s de referencia para el método comparativo, un intervalo de referencia para un nuevo método puede ser calculado a partir de la línea de regresión, o estadística de regresión, obtenida a partir de la información del experimento de comparación de métodos.

Capítulo 16. ¿Cómo evaluar el desempeño de un método?

¿Cómo se incorpora un requisito de la calidad en una Carta de Decisión del Método?

La escala del Cuadro de Decisión de Método está basada en el error total permitido (TE_a) que es definido para el examen. La escala en el eje “y” va desde cero hasta el valor del TE_a y la escala del eje “x” va desde cero hasta el 50 % del requisito de la calidad ($0,5 * TE_a$).

¿Cuál es una fuente común de requisitos de la calidad para un ensayo?

Para los laboratorios de E.U.A., los errores totales permitidos están definidos para aproximadamente 80 exámenes según el criterio CLIA para el desempeño aceptable en esquemas de evaluación de la competencia (PT, por sus siglas en Ingles). Los criterios CLIA son oficialmente publicados en el Registro Federal. Para los laboratorios fuera de Estados Unidos, existen criterios similares disponibles de otros programas nacionales o profesionales de “evaluación externa de la calidad”.

¿Cuáles son los parámetros de desempeño críticos que se grafican en una Carta de Decisión del Método?

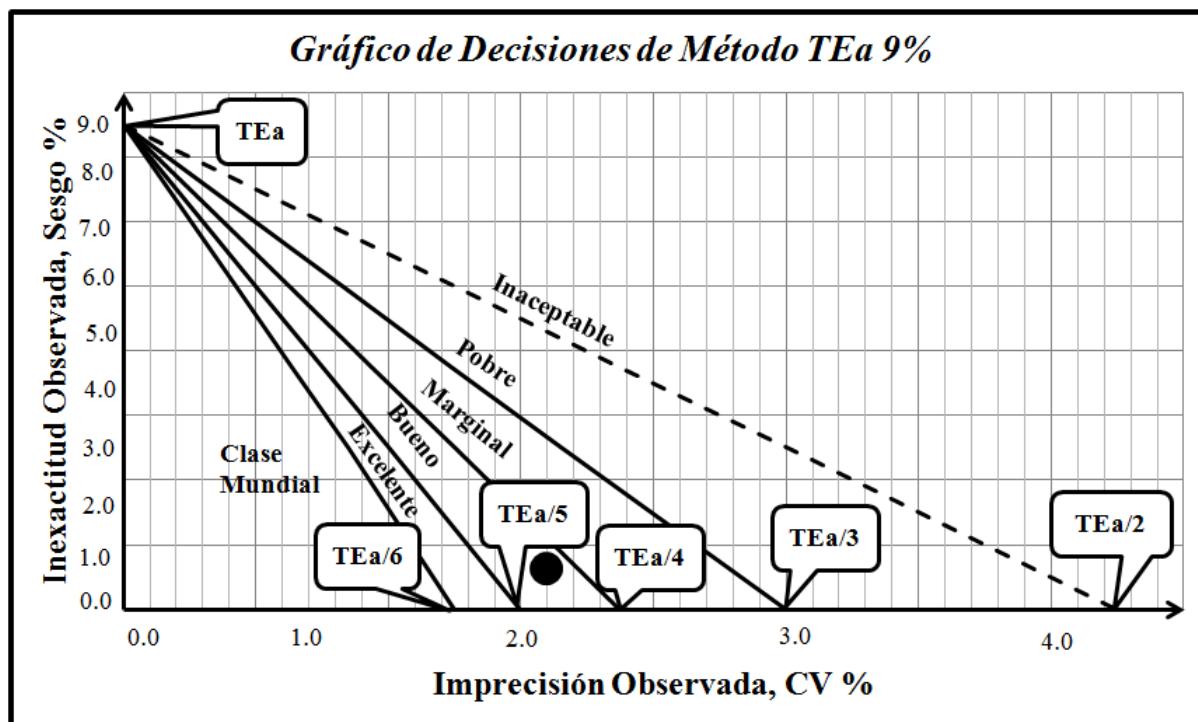
El sesgo o inexactitud observada del método se grafica en el eje “y”. La imprecisión observada o CV del método se grafica en el eje “x”. Se recomienda que el CV del experimento de replicación a largo plazo se acople con el error sistemático (Sesgo) del experimento de comparación de métodos para localizar el “punto operativo o de funcionamiento” del método. Estos resultados brindarán la evaluación más exigente del desempeño del método.

¿Cuáles son las clasificaciones de desempeño en una Carta de Decisión del Método?

Al desempeño del método se lo clasifica como inaceptable, pobre, marginal, bueno, excelente y clase mundial (Seis Sigma), dependiendo del lugar dónde se posicione el punto operativo. Un método con un desempeño inaceptable es tan malo que la calidad del examen no puede satisfacer el requisito de la calidad definido para el examen. Un método con un desempeño pobre podría haber sido considerado aceptable antes de la introducción de la Gestión de la Calidad Seis Sigma, pero Seis Sigma estableció un mínimo de desempeño de 3 sigmas para cualquier proceso de producción. Un método con un desempeño marginal está bien cuando todo está funcionando adecuadamente pero generalmente requerirá esfuerzos extensivos para mantener el desempeño bajo operaciones de rutina. Un método con un desempeño bueno requiere de un esquema de control estadístico interno de la calidad sólido y ofrece un desempeño consistente durante la operación de rutina. Un método con un desempeño excelente será muy manejable y controlable. Un método con desempeño de clase mundial requerirá un mínimo control estadístico interno de la calidad para garantizar que se alcanza la calidad deseada en la operación de rutina.

¿Cuál es su evaluación sobre un método de calcio cuya S observada es 0.2 mg/dL y cuyo Sesgo observado es 0.1 mg/dL a un nivel de decisión médica de 11.0?

Para el calcio, el error total permitido por CLIA es 1,0 mg/dL, el cual a un nivel de decisión de 11,0 mg/dL sería aproximadamente 9%. Establezca una escala en el eje “y” desde cero hasta el valor del requisito de la calidad escogido para el nivel de decisión médica evaluado (9%) en la Carta de Decisión de Método. Establezca la escala en el eje “x” desde cero hasta el 50% del requisito de la calidad escogido, 4,5% (9% *0,5). La S observada es de 0,2 mg/dL a 11,0 mg/dL y equivale a un CV de 1,8%. El sesgo observado es equivalente a 0,9%. El punto operativo ($y=0,9\%$, $x=1,8\%$) muestra que el desempeño del método es “bueno”.



Capítulo 17. ¿Cuál es un procedimiento práctico para validar un método?

¿Cuáles son los experimentos mínimos necesarios para sus exámenes?

La respuesta, por supuesto, depende de la clasificación de los exámenes que se realizan en su laboratorio. Si se realizan sólo exámenes *waived* o PPM, no se necesitan estudios de validación de método. Si se realizan exámenes *non-waived*, por lo general necesitará validar el rango reportable, realizar un experimento de replicación para validar la precisión y realizar un experimento de comparación de métodos para validar la exactitud del método. Puede verificar la transferencia de los intervalos de referencia mediante el cálculo de los resultados de la comparación de métodos si previamente ha realizado estudios para establecer intervalos de referencia en su propio laboratorio o mediante el enfoque de 20 muestras para verificar los intervalos recomendados por el fabricante.

¿Cuál es el número mínimo de mediciones necesarias para cada experimento?

Las mediciones mínimas para el rango reportable son duplicados en una serie de 4 muestras o estándar (o 5 materiales para CAP). La precisión requiere 20 mediciones por un período de 20 días preferiblemente, aunque algunas agencias de inspección pueden aceptar protocolos más cortos. La exactitud a través de la comparación de métodos debería requerir de 20 a 40 especímenes de pacientes – 20 si las muestras son seleccionadas para cubrir un rango analítico amplio, 40 si son seleccionadas de manera

aleatoriedad. La transferencia de los intervalos de referencia puede ser verificada con 20 especímenes de un subgrupo definido de la población.

¿Cuál es la mejor forma de organizar estos experimentos?

Realice la linealidad primero, luego la replicación y la comparación de métodos y por último verifique la transferencia del intervalo de referencia.

¿Cuál es la forma más práctica de hacer los cálculos a partir de los datos?

Esto, por supuesto, depende de sus recursos personales y de los recursos disponibles en su laboratorio. A menudo es práctico usar una hoja de cálculos electrónica ya que la mayoría de las computadoras vienen con algún tipo de paquete de “office” que incluye una hoja de cálculo.

Capítulo 18. ¿Cómo usa la estadística en el Mundo Real?

El análisis de regresión lineal para una prueba de hemoglobina glicosilada muestra lo siguiente: Nuevo Método (Y) = 0.93 Viejo Método (X) + 0.29 ($r = 0.992$). ¿Cuál es el estimado para la pendiente y cuál es el estimado para el intercepto-Y?

Es bastante común ver las estadísticas de regresión presentadas como la ecuación $y=a+b*x$, en la cual “a” es el intercepto-Y y “b” es la pendiente. En este caso “y” es valor del “nuevo método” y “x” es el valor del “viejo método”. El multiplicador del “viejo método” es la pendiente, que es de 0,93. Idealmente, debería ser de 1,00 pero la información muestra que el nuevo método tiene un error proporcional de 7,0% (1,00-0,93, que es 0,07 o 7,0% al expresarlo como porcentaje). El intercepto-Y es de 0,29% GHb. Nótese que % aquí es la unidad de la medición GHb.

¿Son confiables los estimados para la pendiente y el intercepto-y?

Dado que “r” es mayor que 0,99, la regresión lineal común brindará estimados confiables de la pendiente y el intercepto-Y.

¿Debería usted emplear la prueba “t” o Bland Altman en lugar de la regresión lineal simple?

No es necesario recurrir al análisis de datos de una prueba “t” o Bland-Altman ya que la regresión lineal común será confiable para estos datos.

¿Debería usted usar la regresión de Deming o el modelo Passing-Bablock en lugar de la regresión lineal simple?

No es necesario utilizar la regresión Deming o Passing-Bablock ya que la regresión lineal común será confiable para esta información.

Capítulo 19. ¿Cómo pueden ser verificadas las especificaciones de desempeño de los fabricantes?

¿Qué es lo que hace al protocolo EP 15 atractivo para los laboratorios?

El EP15 requiere mínima información, sólo 5 días para la precisión, sólo 20 especímenes de pacientes para la comparación de métodos, y posiblemente sólo 2 materiales de referencia analizados en duplicado por 5 días para la veracidad. El EP15 sólo está destinado a verificar las especificaciones declaradas por los fabricantes, es decir para demostrar que el desempeño observado en el laboratorio es consistente con el desempeño esperado (declarado) por el fabricante.

¿Cuáles son las dificultades potenciales en la implementación del EP 15?

Los cálculos son más complicados, particularmente aquellos empleados en la estimación de la precisión. Los cálculos también requieren la estimación de los límites de confianza o límites de verificación para demostrar que el desempeño del laboratorio es consistente con el desempeño declarado por el fabricante.

¿Qué es la precisión intra laboratorio?

Este es un nuevo término, consistente con la terminología utilizada por ISO (Organización Internacional para la Estandarización) para la precisión total o día a día.

¿Qué es “veracidad”?

Este es un nuevo término, una vez más utilizado por ISO, para la exactitud, la cual es estimada al calcular el “sesgo”.

¿Cuál es el modelo estadístico recomendado por el EP 15 para la estimación del sesgo?

Las estadísticas apareadas de la prueba “t”, junto con un gráfico de diferencias para realizar una presentación gráfica de los datos de pacientes.

Capítulo 20. ¿Cómo pueden ser evaluadas las especificaciones de desempeño declaradas por los fabricantes en una escala Sigma?

¿Cuál es el vínculo que existe entre La Carta de Decisión de Método y la métrica Sigma?

El criterio de la Carta de Decisión de Método corresponde al desempeño en la escala sigma, por ejemplo $TE=sesgo+6S$ corresponde a 6-desempeño Sigma, $TE=sesgo+5S$ corresponde a 5- desempeño Sigma, etc. Puede usar una herramienta gráfica para juzgar la aceptabilidad del desempeño, o puede calcular una métrica Sigma.

¿Cuál es la ecuación para calcular una métrica Sigma?

$\text{Sigma} = (TEa-sesgo)/CV$, en el cual se toma TEa como el error total permitido expresado por el criterio CLIA para el desempeño aceptable de exámenes de evaluación

de la competencia (PT, por sus siglas en inglés), el sesgo es la exactitud (error sistemático, veracidad) determinado por un experimento de comparación de métodos y CV es la precisión determinada por un experimento de replicación. Todas estas cantidades deben expresarse en la misma unidad, sea porcentaje o unidades de concentración.

¿Cómo puede estimar la métrica sigma a partir de las especificaciones de desempeño declaradas por el fabricante?

Los fabricantes deben establecer y declarar especificaciones de desempeño para la precisión y la exactitud de sus procedimientos de medida. La especificación de precisión se presenta frecuentemente como la S o CV observado en 2 o más materiales de control diferentes con diferentes concentraciones. La especificación de desempeño de exactitud se presenta frecuentemente a través de las estadísticas de regresión obtenidas a partir de los datos de la comparación de métodos. Debe interpretar adecuadamente las estadísticas disponibles para elaborar el estimado de precisión y sesgo, luego introducir esos números en la ecuación Sigma. Esto requiere la definición del “nivel(es) de decisión médica” o concentración(es) crítica(s) para el uso clínico e interpretación del resultado del examen con el fin de realizar los estimados apropiados de precisión y sesgo a partir de las especificaciones declaradas por los fabricantes.

¿Cuál es el desempeño Sigma de un método de colesterol para el cual la precisión declarada por el fabricante es de 2.0% a una concentración de 200 mg/dL y la exactitud declarada está dada por la siguiente ecuación de regresión $Y=0.96X+6.0$?

Dado el requisito de la calidad de CLIA TE_a de 10% y un nivel de decisión médica de 200 mg/dL, el CV de 2,0% representa el desempeño a una concentración crítica. El sesgo a 200 mg/dL es estimado como 1,0% ($Y_c=198$ correspondiente a un X_c de 200, o un sesgo de 2,0 mg/dL, el cual es 1,0% en una concentración de 200 mg/dL). Sigma = (10-1)/2 o 4,5.

Capítulo 21. ¿Qué impacto tendrá ISO en la gestión de la calidad analítica?

¿Cuál es el significado de las siglas ISO?

Organización Internacional para la Estandarización, la cual establece estándares globales para el manejo de la calidad incluyendo un documento específico 15189 para laboratorios médicos.

¿A qué concepto se aproxima más “veracidad”: precisión, exactitud o error total?

A exactitud, que es estimada como sesgo, o diferencia promedio entre las mediciones observadas y el valor verdadero o trazable.

¿A qué concepto se aproxima más “incertidumbre”: precisión, exactitud o error total?

El error total representa el máximo error esperado de un método de medición; de la misma manera, la incertidumbre explica la variación total esperada. Las dos características son calculadas de maneras diferentes, pero ambas intentan caracterizar la calidad de una medición o resultado de un examen.

¿Por qué es importante la trazabilidad?

La trazabilidad es la conexión con la verdad o la relación de un valor observado con el valor real. La trazabilidad depende de un sistema de métodos de referencia o materiales de referencia que conectan métodos de campo con la verdad (valores verdaderos).

¿Cuáles son las ventajas que aportan los conceptos de veracidad e incertidumbre?

Estos términos representan el lenguaje global estándar para describir el desempeño de los procedimientos de medición. Están basados en conceptos y procedimientos provenientes del campo de la metrología, es decir la ciencia de las mediciones. Como tales, representan conceptos y terminología aceptada que serán adoptados en todos los campos de medición y en la actualidad están siendo adoptados globalmente en los laboratorios clínicos. La veracidad y la incertidumbre brindan medidas de la calidad que los procesos de exámenes de laboratorio están alcanzando.

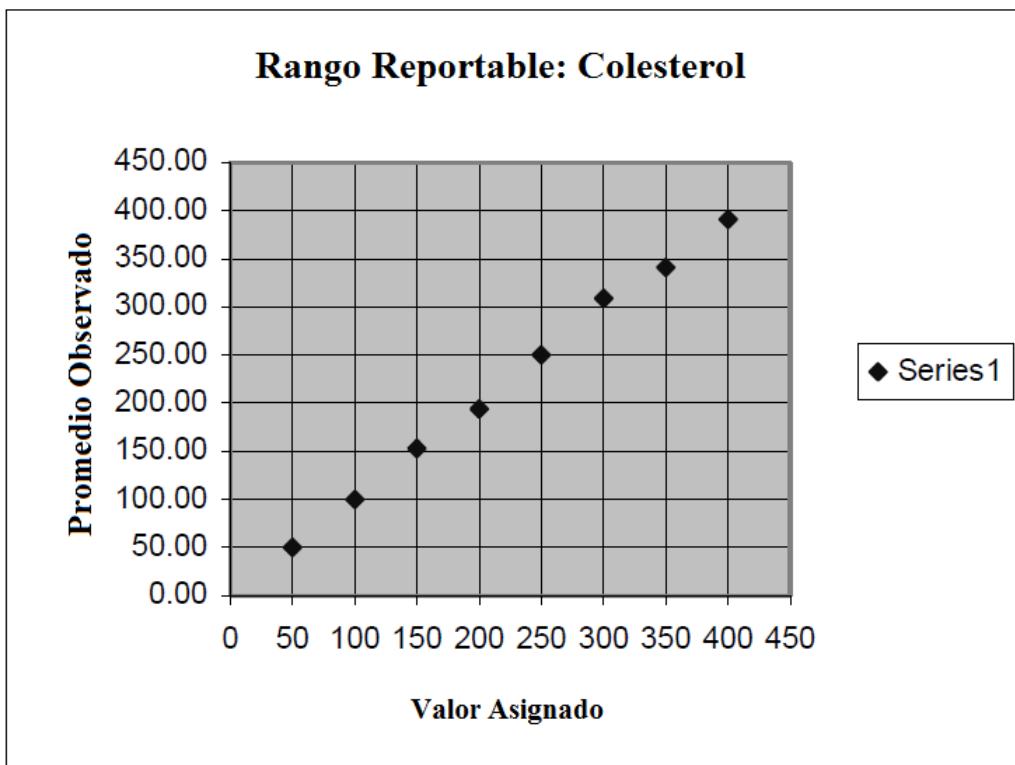
¿Cuáles son las ventajas de aplicar el modelo de error tradicional en el laboratorio clínico?

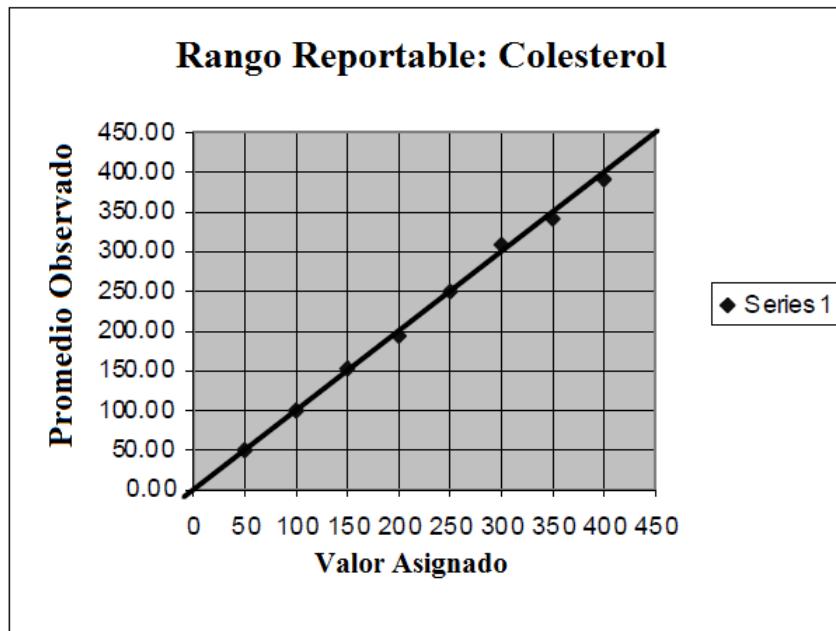
¡La practicidad! Los errores son medidas significativas de la calidad de los exámenes de laboratorio y procesos de medición. La gente entiende que los errores necesitan ser cuidadosamente monitoreados y controlados con el fin de brindar resultados de exámenes con la calidad apropiada. Existen muchas herramientas para medir, monitorear, controlar, validar y evaluar la calidad de los exámenes de laboratorio y sus procesos.

Respuestas al Conjunto de Problemas – Información de Validación del Método de Colesterol

Experimento de Linealidad

Los valores promedio a ser graficados en el eje “y” contra los valores asignados de botella en el eje “x” son los siguientes: 49,50 mg/dL contra 50 mg/dL; 99,75 contra 100; 153,25 contra 150; 195,00 contra 200; 250,75 contra 250; 308,50 contra 300; 341,75 contra 350; 389,75 contra 400. El trazado de una línea recta a través de estos puntos muestra buena linealidad a lo largo del rango. La caída observada a 400 mg/dL es de 10.25 mg/dL, es pequeña comparada con el error permitido del 10% (el criterio CLIA PT “por sus siglas en inglés” para desempeño aceptable), que expresado en unidades de concentración es de 40 mg/dL en el extremo superior ($400\text{mg/dL} \times 0,10$). El rango reportable final ha sido validado como 0 a 400 mg/dL.





Experimento de replicación

Los datos para el Control A dan una media de 199,20 mg/dL, una S de 5,84 mg/dL y un CV de 2,93%. El Control B muestra una media de 239,65 mg/dL, una S de 5,61 mg/dL y un CV de 2,34%.

Experimento de recuperación

La cantidad de estándar añadido es 50,0 mg/dL (dilución 1/10 de un estándar de 500 mg/dL). Para 6 muestras de recuperación, el resultado de base promedio (muestra de base), el resultado “spiked” (la muestra suplementada) promedio, la diferencia y el % de recuperación es el siguiente:

Muestra de Base	Muestra Suplementada	Diferencia	Agregado	% de Recuperación
149,75	200,50	50,75	50,0	101,5
182,75	228,50	45,75	50,0	91,5
201,75	253,25	51,50	50,0	103,0
189,00	240,50	51,50	50,0	103,0
160,50	207,75	47,25	50,0	94,5
190,25	242,25	52,00	50,0	104,0

La recuperación promedio es de 99,58%, la cual es muy cercana a la recuperación ideal de 100,0%. Se concluye que hay poco, si es que existe, error proporcional sistemático observado para este método.

Experimento de interferencia

El efecto de la bilirrubina en el desempeño del método puede ser visto al calcular la diferencia entre las muestras de base promedio y las muestras “spiked” promedio.

Base	Suplementada	Diferencia
214,25	225,50	11,25 mg/dL
220,25	234,75	14,50
295,00	302,00	7,00
171,25	183,50	12,25
251,75	260,00	8,25
225,50	236,00	10,50

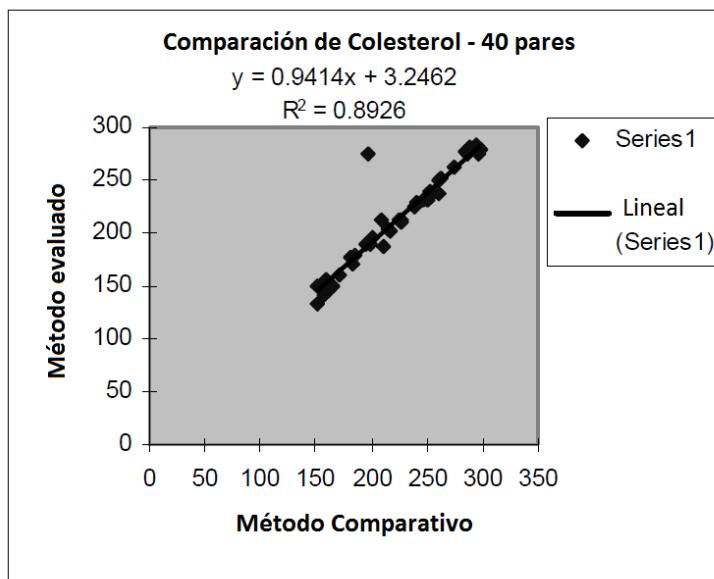
La interferencia promedio es de 10,63 mg/dL cuando el espécimen tiene un nivel de bilirrubina de 10 mg/dL. Se concluye que una bilirrubina elevada es un problema con este método de medición de colesterol.

Experimento de comparación de métodos

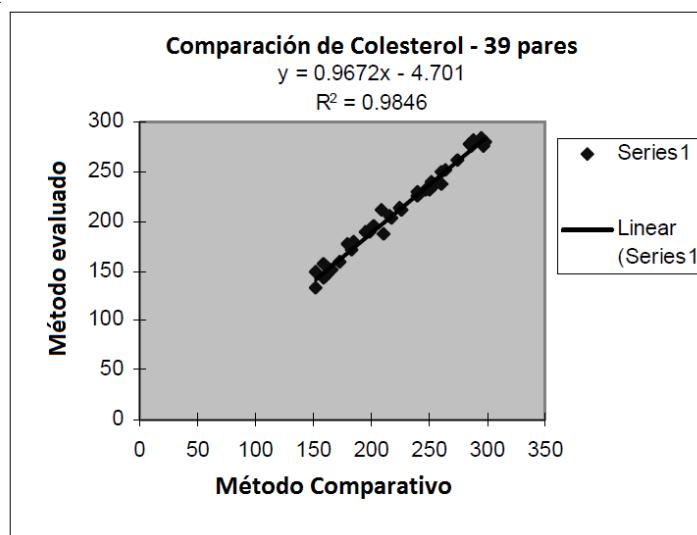
Los datos originales de 40 pares de datos, muestran que hay un par para el cual hubo probablemente un error al registrar los resultados. La muestra #8 muestra resultados tabulados de 197 para el método comparativo y 275 para el método de examen. Incluir este par de resultados fue un truco sucio para enseñarle la importancia de inspeccionar la información cuidadosamente tan pronto como los resultados analíticos estén disponibles.

Si este punto discrepante se incluye en el análisis estadístico de los datos, las estadísticas de regresión dan una pendiente de 0,941, un intercepto-Y de 3,25 mg/d, una S de 15,3 mg/dL para los residuales o puntos sobre la línea de regresión, y un coeficiente de correlación de 0,945. Las estadísticas de la prueba “t” para datos apareados dan un sesgo de -9,73 mg/dL en una media de 211,5 mg/dL, con una S de diferencias (S_{dif}) de 15,4 mg/dL y un valor “t” de 4,00.

Si se saca este punto discrepante del conjunto de datos y se analizan los 39



pares de resultados restantes, las estadísticas de regresión dan una pendiente de 0.967, un intercepto-Y de -4,70 mg/dL, una S de 5,73 mg/dL para los puntos sobre la línea de regresión, y un coeficiente de correlación de 0.992. Las estadísticas de la prueba “t” para datos apareados dan un sesgo de -12,0 mg/dL en una media de 209,9 mg/dL, una S de diferencias (S_{dif}) de 5,86 mg/dL y un valor “t” de 12,8. El punto discrepante ejerce mayor influencia en términos de S y coeficiente de correlación porque está en el medio del rango analítico de los datos. Si hubiera una discrepancia similar cerca del extremo superior o inferior del rango analítico, la pendiente y el intercepto se verían más afectadas.



Para este conjunto de datos, habiendo sacado el punto discrepante, las es-

tadísticas de regresión son la mejor opción para analizar los datos. El coeficiente de correlación alto indica que el rango analítico de los datos es lo suficientemente amplio y los coeficientes de regresión (pendiente e intercepto-Y) deberían ser confiables.

En una concentración que representa un nivel de decisión médica de 200 mg/dL (que sería crítica para la interpretación de un examen de colesterol), el error sistemático puede ser calculado a partir de la línea de regresión, como se muestra a continuación: $Y_c = a + bX_c$, en el cual " X_c " es 200 mg/dL, "a" es -4,70 mg/dL, y "b" es 0,967. El valor correspondiente a " Y_c " sería de 188,7 mg/dL, lo que indica un error sistemático de 11,3 mg/dL o 5,65%.

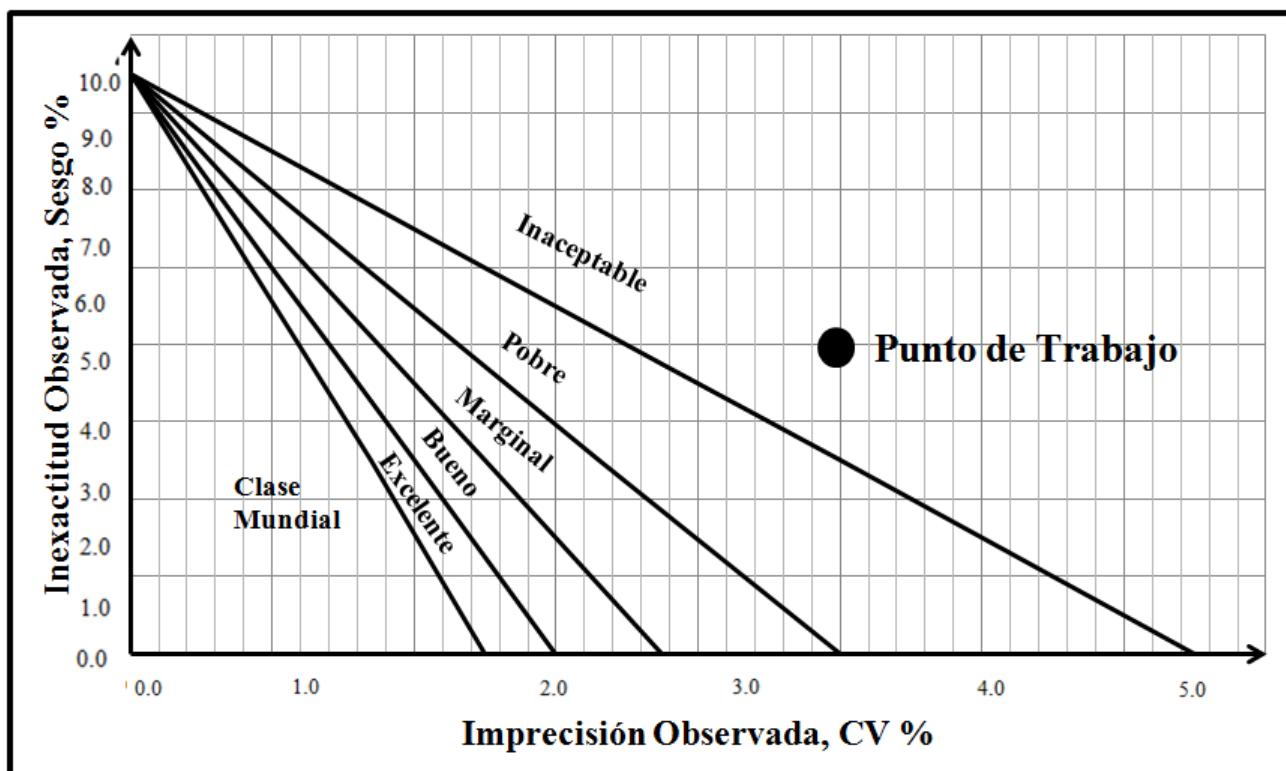
Conjunto de Problemas de Límite de Detección

El límite de Blanco (LoB) puede ser calculado a partir del estándar "0" como $1,65 * 0,26 = 0,43$. El Límite de Detección (LoD) puede ser calculado de la S del estándar "2" más el LoB calculado previamente, es decir $LoB + 1,65 * 0,56 = 1,37$. La sensibilidad funcional se calcula más fácilmente que el Límite de Cuantificación (LoQ). El CV del estándar "6" es muy cercano a 20%, la FS es de aproximadamente 6,0. El LoQ será más alto, pero es difícil de estimar a partir de estos datos.

	Est 0	Est 2	Est 4	Est 6	Est 8
	-0,23	2,84	5,51	6,46	6,92
	0,12	2,26	5,07	5,14	6,29
	-0,40	1,25	3,15	3,78	8,67
	0,13	3,52	2,72	7,28	7,86
	0,05	1,95	3,67	6,20	5,24
	0,18	2,71	5,35	5,00	7,38
	-0,41	3,07	5,22	5,54	7,41
	0,20	2,47	4,46	6,44	6,06
	-0,32	1,95	2,55	6,71	7,53
	-0,08	1,94	3,61	7,17	10,11
	-0,41	1,93	3,79	4,92	8,70
	-0,39	2,45	3,21	4,40	6,37
	-0,32	2,38	3,42	6,11	7,54
	-0,37	1,37	4,73	7,51	9,30
	-0,17	1,58	4,07	6,38	8,10
	0,38	2,46	3,47	5,55	7,87
	-0,31	2,55	3,28	8,19	8,20
	-0,41	2,57	5,67	8,48	7,99
	0,10	2,27	4,99	5,45	9,38
	-0,30	1,57	4,42	7,61	8,22
	Est 0	Est 2	Est 4	Est 6	Est 8
media	-0,15	2,25	4,12	6,22	7,76
DS	0,26	0,57	0,97	1,26	1,20
CV	-173,50	25,44	23,46	20,28	15,44

Evaluación del desempeño del método

Se puede confeccionar una Carta de Decisión de Método usando esta información para un requisito de la calidad analítica del 10%, que es el que dicta CLIA como el error total permitido (TEa) para una prueba de colesterol. El punto operativo está representado por una coordenada “x” de 2,93% que sale de los datos para el Control A (cuya media es cercana a 200 mg/dL) del experimento de replicación. La coordenada “y” es de 5,65% en base al error sistemático observado a un nivel de 200 mg/dL proveniente de los datos de la comparación de métodos. La Carta de Decisión de Método clasifica este método con un desempeño inaceptable, por lo tanto no debería considerar el uso de este método para exámenes de rutina en su laboratorio.



Apéndice 1: Calidad Analítica CLIA '88 Requisitos

Es fundamental para el Control de la Calidad conocer la calidad que se necesita controlar. Se debe definir un requisito de la calidad si se desea que la práctica del Control de la Calidad sea racional y objetiva. Históricamente, el Control de la Calidad se ha utilizado en los laboratorios para monitorear cualquier calidad alcanzada, y no para controlar un nivel definido de la calidad.

Requisitos de la Calidad Analítica

Los requisitos de la calidad analítica han sido definidos por los criterios de Evaluación de la Competencia CLIA '88 para un desempeño aceptable [Registro Federal, 28 de Febrero, 1992;57(40):7002–186].

Estos criterios se presentan en tres formas distintas:

- como límites de concentración absoluta, por ejemplo, valor target ± 1 mg/dL para calcio.
- como porcentaje, por ejemplo, valor target $\pm 10\%$ para albúmina, colesterol, y proteínas totales.
- como la distribución de una encuesta de evaluación externa de la calidad para un grupo de laboratorios (grupo par o por método), por ejemplo, valor target ± 3 desvíos estándar (SD) para la hormona estimulante de la tiroides.

En unos pocos casos, se presentan dos tipos de límites, por ejemplo, el requisito para glucosa se da como el valor target ± 6 mg/dL o $\pm 10\%$ (cuálquiera que sea mayor).

El criterio CLIA de Evaluación de la Competencia especifica los errores totales permitidos. El formato de error total es absoluto dado que las reglas CLIA-88 especifican que solo una prueba sea realizada para cada muestra de Evaluación de la Competencia.

Bajo tales condiciones, el error analítico observado será el error total debido tanto a la imprecisión como a la inexactitud.

A continuación se muestra un listado de los criterios de Evaluación de la Competencia CLIA para las pruebas actualmente reguladas, en base al Registro Federal, 28 de Febrero, 1992.

Química de Rutina

Prueba o Analito	Desempeño Aceptable
Ácido úrico	Valor target ± 17%
Alanina aminotransferasa (ALT)	Valor target ± 20%
Albúmina	Valor target ± 10%
Amilasa	Valor target ± 30%
Aspartato aminotransferasa (AST)	Valor target ± 20%
Bilirrubina Total	Valor target ± 0,4 mg/dL o ± 20 % (el mayor)
Calcio total	Valor target ± 1,0 mg/dL
Cloro	Valor target ± 5%
Colesterol lipoproteína de alta densidad (C-HDL)	Valor target ± 30%
Colesterol total	Valor target ± 10%
Creatina kinasa	Valor target ± 30%
Creatina kinasa isoenzimas	MB elevada (presente o ausente) o valor target ± 3 SD
Creatinina	Valor target ± 0,3 mg/dL o ± 15% (el mayor)
Fosfatasa alcalina	Valor target ± 30%
Gases en sangre pCO2	Valor target ± 5 mm Hg o ± 8% (el mayor)
Gases en sangre pH	Valor target ± 0,04
Gases en sangre pO2	Valor target ± 3 SD
Glucosa	Valor target ± 6 mg/dL o ± 10%
Hierro total	Valor target ± 20%
Lactato deshidrogenasa (LDH)	Valor target ± 20%
LDH isoenzimas	LDH1/LDH2 (+ o -) o valor targer ± 30%
Magnesio	Valor target ± 25%
Nitrógeno ureico	Valor target ± 2 mg/dL o ± 9% (el mayor)
Potasio	Valor target ± 0,5 mmol/L
Proteínas totales	Valor target ± 10%
Sodio	Valor target ± 4 mmol/L
Triglicéridos	Valor target ± 25%

Toxicología

Prueba o Analito

Prueba o Analito	Desempeño Aceptable
Ácido valproico	Valor target ± 25%
Alcohol, sangre	Valor target ± 25%
Carbamazepina	Valor target ± 25%
Digoxina	Valor target ± 20% o 0,2 ng/mL (el mayor)
Etosuccimida	Valor target ± 20%
Fenitoína	Valor target ± 25%
Fenobarbital	Valor target ± 20%
Gentamicina	Valor target ± 25%
Litio	Valor target ± 0,3 mmol/L o ± 20% (el mayor)
Plomo, sangre	Valor target ± 10% o ± 4 mcg/dL (el mayor)
Primidona	Valor target ± 25%
Procainamida (y metabolito)	Valor target ± 25%
Quinidina	Valor target ± 25%
Teofilina	Valor target ± 25%
Tobramicina	Valor target ± 25%

Hematología

Prueba o Analito

Prueba o Analito	Desempeño Aceptable
Identificación celular	90% o mayor acuerdo en identificación
Diferencial glóbulos blancos	Target ± 3 SD en base al porcentaje de diferentes tipos de glóbulos blancos
Fibrinógeno	Target ± 20%
Hematocrito	Target ± 6%
Hemoglobina	Target ± 7%
Recuento de eritrocitos	Target ± 6%
Recuento de leucocitos	Target ± 15%
Recuento de plaquetas	Target ± 25%
Tiempo de protrombina	Target ± 15%
Tiempo de tromboplastina parcial	Target ± 15%

Endocrinología

Prueba o Analito

Prueba o Analito	Desempeño Aceptable
Cortisol	Valor target ± 25%
Gonadotrofina coriónica humana	Valor target ± 3 SD o (positivo o negativo)
Hormona estimuladora de la tiroides	Valor target ± 3 SD
T3 captación	Valor target ± 3 SD por método
Tiroxina	Valor target ± 20% o 1.0 mcg/dL (el mayor)
Tiroxina libre	Valor target ± 3 SD
Triiodotironina	Valor target ± 3 SD

Inmunología General

Prueba o Analito

Prueba o Analito	Desempeño Aceptable
Alfa-1-antitripsina	Valor target ± 3 SD
Alfa-fetoproteína	Valor target ± 3 SD
Anticuerpos anti-nucleo	Valor target ± 2 diluciones o (pos. o neg.)
Antiestreptolisina O	Valor target ± 2 diluciones o (pos. o neg.)
Anti-virus de inmunodeficiencia humana	Reactivo o no reactivo
Complemento C3	Valor target ± 3 SD
Complemento C4	Valor target ± 3 SD
Factor reumatoideo	Valor target ± 2 diluciones o (pos. o neg.)
Hepatitis (HBsAg, anti-HBc, HBeAg)	Rectivo (positivo) o no reactivo (negativo)
IgA	Valor target ± 3 SD
IgE	Valor target ± 3 SD
IgG	Valor target ± 25%
IgM	Valor target ± 3 SD
Mononucleosis infecciosa	Valor target ± 2 diluciones o (pos. o neg.)
Rubeola	Valor target ± 2 diluciones o (pos. o neg.)

Índice

A

AC, ver aseguramiento de la calidad
aceptabilidad del método
72, 83, 84, 132, 133, 153, 166, 187, 225, 228, 253
análisis de regresión 89, 100, 102, 222, 224, 227, 302
análisis de varianzas (ANOVA) 122
análisis estadístico 82, 144, 189, 192, 223, 272, 308
aplicaciones de campo de la validación de métodos 197– 204
aseguramiento de la calidad 5 , 40

B

b, ver pendiente
buenas prácticas de laboratorio 175

C

calculador de datos apareados 79, 82, 242
Calculador de Desvío Estándar (S) 79, 80, 82
calibración 41, 44-46, 49, 57, 59, 67, 69, 71, 103, 164, 225, 234, 235, 265, 266, 268, 269, 275, 276, 290
definición 107
directivas de los fabricantes 105
material 109
puntos múltiples 105
punto de ajuste 105
dos puntos 106
punto cero 106
verificación de la calibración 107, 112
análisis de datos 110
calidad 3, 5-7, 15
definición ISO 265
calidad analítica 5-7, 16, 17
características de aplicación 57-61, 225, 288
características de desempeño 6, 37, 43, 57-61, 231, 234, 265, 266, 269, 278, 277, 288, 289
características del método 209

características metodológicas 57-61, 65, 213
carta de decisión de métodos 195, 196, 199, 201, 204, 211, 213, 300
calculador 84
ejemplos de aplicación 202
desempeño excelente 201
desempeño bueno 201
como construir 199-200
como usar 201
desempeño marginal 201
desempeño pobre 201
desempeño inaceptable 201
desempeño clase mundial 201

CC, ver control de la calidad
CLIA, CLIA criterios de aceptación, etc. 15, 22-24, 33, 39-40, 42-44, 47-49, 51, 70, 81, 105-107, 111-112, 121, 133-136, 163, 167, 179, 189, 198, 199, 202, 203, 209, 211, 233, 251, 257, 258, 260, 269, 287, 300, 303, 304, 306, 312, 313

Clinical Laboratory Improvement Amendments, ver CLIA

Clinical Laboratory Standards Institute, ver CLSI

CLSI 24, 70, 73, 108-110, 113, 118, 122-123, 135-136, 161, 175, 179-180, 187-188, 190, 193, 231, 233-234, 237, 248, 265, 276, 279-280, 299

CMS 7, 24, 37, 39, 42, 46, 49
coeficiente de correlación (r) 30-31, 82-83, 102, 135, 148-149, 151, 155, 212, 287, 291, 297-298, 308-310
definición 102

coeficiente de Pearson, ver coeficiente de correlación
coeficiente de variación (CV) 32, 76, 90, 102, 122, 124, 199, 212, 238, 257, 259, 280, 291
definición 90

COLA 24, 39, 50
College of American Pathologists (CAP), 19, 37, 39, 107

complejidad del examen 42
concentración de decisión médica (nivel de decisión médica)
102, 119, 126, 132, 151, 153, 155, 165, 196, 198-199, 213, 226, 245, 259, 296, 216

Conferencia de Consenso de Estocolmo	19	desviación estándar de las diferencias	228, 241
control arbitrario	22	desviación estándar total	121
control de la calidad	6, 9, 41, 46-48, 51, 60, 72, 119, 135, 201, 203, 235, 265, 267, 269, 285, 313	desvío sistemático	34
control estadístico interno de la calidad	6, 8-9, 18, 20-22, 56, 73, 112, 241, 257, 269, 271, 279, 300	detección de errores	6
control externo de la calidad	271	diferencia promedio entre métodos	160, 163
corrida analítica	201-202, 235	diseño de la calidad	10
covarianzas	273	dispersión	90
criterios de aceptación de los esquemas de evaluación de la competencia	20-21, 46, 60-6, 81, 109, 112	distribución	31, 79
cuadrados mínimos, línea	100, 132, 287	distribución chi-cuadrado	239-240
curva Gaussian	90		
curva normal	90		
CV ver coeficiente de variación			
CV máximo permitido	22		
D			
datos aberrantes	227	ecuación de la línea de regresión	101
desempeño aceptable	22, 24, 34, 133-134, 167, 179, 202-203, 258, 260, 278, 300, 303, 313	ecuación de regresión 100-101, 152, 192, 252, 304	
decisión de métodos	133, 253	EP5	73, 122, 248
decisión sobre el desempeño del método	197-204	EP6	113
defectos por millón (DPM)	254-255	EP9	73, 136, 248
desempeño del método	6, 10, 22-23, 33, 48, 56, 60, 65, 67, 71-72, 78, 87, 133, 154, 179, 187, 195, 197, 209, 211, 213-214, 216, 225, 235, 266, 278, 286, 287, 289, 293, 300, 308	EP15	73, 122, 123, 135-136, 248, 265, 282, 303
desviación estándar (SD)	18, 30-31, 79, 82, 87-94, 101, 111-112, 117, 120-122, 124, 133, 144, 151, 176-177, 199, 212, 228, 237, 241, 243, 274, 287, 290, 292-293, 299	cálculos usando Excel	237
desviación estándar del intercepto (sa):	101	cálculos de datos de precisión	236
desviación estándar de la pendiente (sb):	101	Protocolo de precisión	235-244
desviación estándar sobre la línea de regresión	100	propósito	233
desviación estándar de los residuales	100	alcance y definiciones	234
		hoja de cálculo	237
		cálculo de datos de veracidad	246
		protocolo de veracidad empleando datos de pacientes	241-244
		protocolo de veracidad empleando materiales de referencia	245-247
		EP17	175, 180-181
		error aleatorio	31-32, 66-68, 78, 80, 82-83, 91, 95-98, 110-112, 117-120, 124, 133, 135, 140, 142, 144-145, 148-149, 151, 153, 160, 165, 198-199, 210, 213, 225, 228, 272, 286-287, 296
		EA, ver error aleatorio	
		errores analíticos	7-8, 10, 27-28, 30-31, 33, 63, 65, 67, 69, 72, 89, 100, 102, 131, 139, 141, 144, 150-151, 154, 198, 210, 211, 282, 296, 298
		definiciones	31
		error constante	31, 151, 159, 290, 296-297
		error estándar de la media	89, 92, 95, 246
		definición	89

error máximo permitido (ETa), ver también error total máximo permitido	268	estadísticas de regresión	79, 80, 82, 102, 125, 132, 137, 140, 141, 151, 152, 155, 162, 191, 192, 199, 212, 213, 216, 224, 226, 227, 290, 291, 295, 296, 297, 302, 304, 308, 309
error observador	167, 225	estadística no paramétrica	91
errores pos analíticos	7, 10	estadística paramétrica	91
errores pre analíticos	7, 10	estadísticas de regresión lineal	82, 132, 212, 290, 295-296
error total y error total permitido	19, 21, 23, 32-34, 58-59, 65, 81, 110-112, 121, 133-134, 155, 167, 176, 197-199, 202-203, 211, 213, 225, 264, 272, 275-281, 286-287, 2929, 299-300, 303, 305, 312-313	estadísticamente significativo	84, 88, 96, 134, 140, 144, 153-154, 292, 296
ET ver error total		estado del arte	20, 22, 278
ETA ver error total permitidor		estándares de la calidad	17-19, 20, 23, 211
ET calculado	133	una corta historia	18-20
ES, ver también error sistemático		fuentes convenientes	22
Error Sistemático (ES)	32, 133, 135	jerarquía	19-20, 268, 278
error sistemático constante	32, 67-68, 82, 140, 142, 144, 146, 150-152, 157, 167, 171, 212, 296-297	tendencias y dirección	24
error sistemático proporcional	32, 67, 67, 69, 82, 93, 140, 142, 144, 147-148, 150-152, 157, 163, 166-168, 170, 212, 225, 290, 296-297	estudio de comparación de métodos	30, 81, 129, 151, 162, 191, 215, 228, 258, 289, 295
escala Sigma	251, 253-254, 257, 303	estadísticas apropiadas	131, 141
especificaciones declaradas por los fabricantes	105, 134, 247, 303-304	método comparativo	72, 98, 127, 128, 157, 162, 192, 299, 308
especificidad analítica	33, 47, 49, 57, 209-210, 269, 275, 288	criterio de desempeño aceptable	260
especímenes	55, 57, 69, 71-72, 79, 81, 83-84, 107, 109, 118, 121, 126, 128-131, 133, 135, 142, 150, 160-163, 165, 167, 170-171, 186-190, 210, 212, 214-215, 226-228, 234, 258, 289, 295, 299, 301-302	estudios mínimos recomendados	122
especímenes de pacientes	69, 83-84, 107, 109, 121, 128, 129, 135, 160-161, 165, 170-171, 186, 210, 212, 234, 258, 289, 299, 301	periodo de tiempo	118
esquemas de evaluación de la competencia		verificación de las especificaciones del abricante	134-135
	19, 46, 60-61, 81, 109, 112, 135, 198, 211, 255, 257, 271-272, 278-279, 300	estudio de interferencias	159, 210
estadísticas	10, 17, 29, 31, 75, 79, 80, 82, 83, 86, 87, 89, 94, 97, 101, 102, 125, 129, 131, 132, 133, 136, 137, 138, 139, 140, 141, 144, 145, 148, 149, 150, 151, 152, 155, 162, 190, 191, 192, 195, 197, 199, 212, 213, 216, 221, 223, 224, 225, 226, 227, 228, 229, 247, 279, 287, 290, 291, 292, 294, 295, 296, 297, 302, 303, 304, 308, 309	método comparativo	162
		cálculos de datos	162
		concentración del material interferente	159, 161
		criterios para desempeño aceptable	167
		solución interferente	160
		interferencias a ser examinadas	161
		desempeño del pipeteo	161
		propósito	159
		replicados	160
		volumen del interferente añadido	160
		estudio del límite de detección	175
		muestra de blanco	176
		solución de blanco	178
		definición	176
		ejemplos estimados	179
		factores a considerar	177
		especificaciones del fabricante	175
		número de mediciones replicadas	178
		propósito	175
		cantidad a ser estimada	179
		muestra suplementada	178
		definición	176

cometarios	180	estudios de replicación finales	71
periodo de tiempo del estudio	179	estudios finales	68
verificación de las especificaciones del fabricante	175	estudio de replicación intra corrida	118
estándar de cero	178	estudios de validación de métodos preliminares	
estudio de linealidad	43-44, 79, 81, 103, 105, 108-109, 131, 152, 158	estudios de validación	27, 64-65, 72, 88, 117, 139, 207, 224, 257, 266, 286-287, 289, 292
análisis de datos	110	evaluación de errores	29, 55, 66, 141
diluyente	109	evaluación de la calidad	5, 7, 16, 40-41, 257, 261
materiales	109	evaluación de la competencia	19-21, 23, 46, 49, 60-61, 81, 109, 112, 135, 198, 211, 234, 255, 257, 271-272, 278-279, 300, 313
cantidad de niveles	109	exactitud	7, 9, 15, 20, 33, 42-43, 46-49, 57-59, 65, 69, 80, 88, 107, 127-128, 134-135, 159, 161, 164, 209, 225, 251-252, 258, 264-265, 269-270, 272-273, 275-277, 280, 287-289, 292-293, 31, 303-305
cantidad de replicados	110	exactitud de la medición, definición ISO	270
procedimiento de dilución	109	expresión de la incertidumbre de medida	272-273
estudio de rango reportable	105-112		
ver también estudio de linealidad gráfico	111		
estudio de replicación	117, 129, 213-215, 287, 294-295		
concentraciones a ser ensayadas	119		
criterio para desempeño aceptable	121		
cálculo de datos	120		
consideraciones futuras	122		
matriz de la muestra	118		
cantidad de materiales	119		
cantidad de muestras a ser ensayadas	120		
estudios mínimos recomendados	122		
periodo de tiempo	118		
verificación de las especificaciones declaradas por el fabricante	121		
intra día	118, 119, 121, 123		
intra corrida	118, 119, 121, 123		
hoja de cálculo	204, 208		
estudio de recuperación	69, 92, 158, 163-167, 167-168, 212, 298, 307		
concentración necesaria del analito	165		
concentración de la solución estándar	165		
calcúlo de datos	166		
factores a considerar	164-166		
número de pacientes analizado	165	gestión de la calidad	14-17, 24, 40, 49, 201, 209, 229, 253-254, 263, 265, 279, 281, 286, 300, 304,
número de mediciones replicadas	165	gestión total de la calidad (TQM)	13, 17, 24, 253, 285
exactitud de pipeteo	164	grados de libertad (df)	90, 92, 95-96, 98, 101, 122, 134, 236, 239-240, 243, 246, 249, 292-293
propósito	163	gráfico de Bland-Altman, ver también gráfico de diferencias	224, 228, 302
comentarios	167-168	gráfico de comparación	79-80, 82-83, 130-131, 135, 139, 145, 154, 212, 214, 226-227, 290, 295
verificación de la técnica experimental	166	gráfico de diferencias	79-80, 82, 84, 130, 135, 139, 154 224, 226-228, 242-243, 247, 303
volumen de estándar adicionado	164		
estudios finales de validación de métodos	70		

guía para la expresión de la incertidumbre de medida (GUM)	265, 270, 272, 276-277	intervalo de referencia	18, 43, 48, 185-190, 192, 295, 299
guía para verificar las especificaciones declaradas por los fabricantes	231, 233	datos demográficos	189
		establecimiento	187
		límites	188
H		inexactitud	6, 21-22, 33, 64, 80, 84, 125, 127, 157, 200-201, 203, 210, 212-213, 285-286, 289, 295, 300
hemólisis	161, 210	Ver también exactitud, sesgo	
herramientas para el análisis de datos	136, 279	inexactitud permitida	84
herramienta para graficar datos de linealidad	79, 81	inexactitud observada	203, 300
herramientas recomendadas para el análisis de datos	78-80	intercepto	100, 297-298, 309
hipótesis nula	94, 98	interferencias	48, 93, 128, 135, 158, 160-161, 210, 228, 266-267, 287, 298
histograma	79-80, 82, 121, 124, 190, 290	interpretación de la prueba-F	87, 98
I		intervalo de confianza	34, 91-94, 134, 140, 243, 246, 274, 281
implementación del método	17, 70	guía para la validación de métodos y control estadístico interno de la calidad	265-268
imprecisión	6, 16, 20-22, 31, 33, 43-44, 58, 64, 66-68, 71, 78, 80, 82-84, 97, 115-119, 121, 124, 160, 165, 167, 200-201, 203, 210, 213, 216, 228, 251, 258, 277, 278, 280, 285-286, 289, 292, 294-295, 313	ISO 15189	34-35, 263, 265, 269, 279, 283
imprecisión a corto plazo	122	calibración de sistemas	268
imprecisión “intracorrida” e “intradía”	119	validación y verificación	265-268
imprecisión observada	201, 203, 292		
imprecisión permitida	84	J	
incertidumbre	34-35, 95-96, 246-247, 263, 265-267, 269-270, 272-282, 305	jerarquía de las especificaciones de la calidad	278
incertidumbre de la medición	246, 265, 267, 272		
definición ISO	272	L	
incertidumbre estándar	273-274	laboratorios de referencia	272
definición ISO	273	límites de control aceptables	45-46
incertidumbre estándar combinada	273-274	límite de detección biológico	175
definición de ISO	273	límite de blanco (LoB)	173-176, 182, 299, 310
incertidumbre expandida	34, 273-274, 278	definición	176
definición ISO	273	límite de detección (LoD)	7, 33, 44, 48, 57, 69, 71, 107, 109, 173-178, 180, 182-183, 210, 212, 266, 269, 288, 298-299, 310
incertidumbre tipo A	272	definición	176
definición ISO	272	límite de cuantificación (LoQ)	173, 175-176, 299, 310
incertidumbre tipo B	272	definición	176
definición ISO	272	límites de tolerancia	254
inmunoensayos	175, 192	línea de mejor ajuste	81-82, 132, 294

línea de regresión	101, 108, 132, 146-147, 191, 291	métodos <i>non-waived</i> no aprobados por la FDA 43, 49, 288	
linealidad	43-44, 79, 81, 103, 105, 108-109, 131, 152, 159, 210, 215, 227, 294, 302, 306	Métrica Sigma	251-256, 258, 260, 304
línea de mejor ajuste	108	mitos modernos de la calidad	5-8
liofilización	119	muestra de blanco	176, 298
lipemia	59, 71, 160-161, 168, 298	muestras de pacientes	72, 88, 101, 107, 112, 119, 127-128, 135, 165, 178, 241, 247, 293
líquido cefalorraquídeo	118	muestras de esquemas de evaluación de la competencia	46
M		muestra suplementada	179-180, 299
mantenimiento preventivo	45-46, 56, 72, 201, 213	N	
Materiales de Referencia Certificados	135	<i>NCEP</i>	120, 203, 205, 260
materiales de control	46, 48, 119, 122, 124, 135, 210, 214, 235, 271, 304	niveles de decisión	120, 126-127, 131-132, 135, 151, 155, 197, 211, 226-227, 235, 245, 257- 259, 295
materiales de control de la calidad	48, 135	nivel de decisión crítico	132
materiales de referencia	45, 135, 234, 245, 247, 268, 270-271, 303	ver también nivel de decisión médica	
matriz	71, 83, 109, 118-119, 128, 163-164, 178, 212, 235, 289	nivel de detección mínimo 1	75
media	32, 68, 81-82, 89-96, 102, 110-112, 120, 132, 134-135, 150-151, 153, 155, 179, 212-213, 216, 224, 226-228, 236-237, 241, 245-246, 270, 291, 296, 307-308, 312	no linealidad	227
definición	90	<i>non-waived</i>	38-39, 41-43, 47, 49, 70, 175
media verdadera	91	número de mediciones	82-83, 89, 91, 86, 98, 122, 215
medición de la incertidumbre	266	P	
mediciones replicadas	91, 112, 286, 294	PCL, ver procesos de la calidad del laboratorio	
mejor línea de ajuste	30, 224	PC, ver planificación de la calidad	
MC, ver mejora de la calidad		pendiente	30, 79, 82, 87, 90, 100-102, 132, 140, 144-145, 147-148, 150-153, 191, 212, 222, 226-227, 257-258, 287, 290-291, 296-298, 302, 308-310
mejora de la calidad (MC)	17	plan de validación	213-214
mensurando	34, 234, 266-267, 270, 272-273	plan experimental	63, 65, 127, 131, 159
definición ISO	270	planificación de la Calidad (PC)	6, 10, 21, 279-280, 286
metas biológicas	20-21, 23, 278, 280	analítica	21, 279
metas de la calidad	6, 20	período de familiarización	235
método comparativo	72, 98, 157, 162, 192, 308	<i>point-of-care</i>	56, 59
método definitivo	271	pools de pacientes	119, 160
método del fabricante	211	precisión	9, 18, 20, 47-48, 68-70, 73, 88, 122-123, 148, 161, 176, 179, 209, 231-235, 237, 240-241, 247, 251-253, 257-260, 265, 267, 269, 276-277, 280, 287, 289, 292-293, 301, 303-304
método de referencia	45, 58-59, 127, 142, 224		
métodos <i>non-waived</i> aprobados por la FDA	42, 47, 49, 288		

ver también imprecisión	
ver también error aleatorio	
ver también coeficiente de variación	
probabilidad (p)	6, 71, 92-93, 96, 176, 213, 274
procedimiento, definición	40
procedimientos de CC	18, 22, 70, 72, 201, 235, 266, 279, 285
procedimientos de examen	38, 47, 265, 267, 276
definición ISO	265
procesos de la calidad del laboratorio	16, 285, 286
proceso estándar para la gestión de la calidad	16
proceso de validación	27, 55-56, 187-188, 209, 213-215
procedimiento operativo estándar	23
programa de acreditación de laboratorios	48
programas de evaluación externa de la calidad (PEEC)	19-20, 22, 60, 198, 272, 278-279
procesos de laboratorio	254, 256
protocolo de operación	71
pruebas de alta complejidad	47
pruebas de complejidad moderada	47
pruebas <i>non-waived</i> modificados o desarrollados en el laboratorio	43
prueba "t"	82-83, 148, 222, 226-228, 231, 241-242, 247, 297, 302-303, 308-309
pruebas <i>waived</i>	42, 48, 287, 301
punto operativo	84, 202-203, 253, 300, 312
Q	
<i>Q-Probe</i>	254-255
R	
rango analítico	71, 78, 107, 110, 115, 131-132, 149, 162, 178, 224, 226, 289, 295, 301, 309-310
rango analítico de medición (AMR)	107
rango de referencia declarado por el fabricante	209
rango lineal	81, 83, 106, 294
rango reportable	42-43, 45-48, 59, 64, 71, 78, 80, 103, 105, 107-109, 111-112, 114, 128, 135, 214-215, 245, 269, 287, 289, 293-294
definición	107
rangos reportables clínicos (CRR)	107
rango reportable y verificación de la calibración	45-46, 107, 112
rango de trabajo	71
reactivos	17, 41, 45-47, 49, 58-60, 71, 105, 112, 122, 209, 213, 234, 268, 286
recuperación	33, 44, 57, 68-71, 91-93, 128, 157- 159, 163-168, 170, 183, 210, 212-213, 215, 267, 287-289, 298, 307-308
recuperación observada	166
Regla Final de CLIA	39-40, 42, 44, 49, 51, 107, 189, 287
reglas de control	10, 60, 215, 280
Regresión de Deming	83, 153, 155, 221-223, 227, 302
regresión lineal simple, ver regresión lineal	
Regresión de Passing-Bablock	83, 153, 155, 302
técnica	227
repetibilidad	68, 179, 212, 214, 234, 237-240, 260, 266-267, 269
definición	234
requisitos de la calidad analítica	313
requisitos de la calidad	1, 6, 10, 18, 41, 196-197, 201, 203, 254, 256-258, 279-280, 300, 313
residuales	227, 308
S	
S_{meas}	ver también imprecisión, coeficiente de variación, CV, precisión
sangre entera	59
SD entre corridas	238
SD intra laboratorio	237-238
SD, ver también desviación estándar	18, 30-31, 79, 82, 87-94, 101, 111-112, 117, 120-122, 124, 133, 144, 151, 176-177, 199, 212, 228, 237, 241, 243, 274, 287, 290, 292, 293, 299
SD (desviación estándar) permitida	18
SD de las diferencias	30, 228, 241, 290

Seis Sigma	253-254, 260-261, 269, 279, 300
una introducción terriblemente corta	253
dos modelos para medir el desempeño de un proceso	254
ejemplo de cálculo de una métrica sigma	258
guía general para el cálculo	259
inspeccionar salidas	254
poder	260
evaluación de la Calidad de un proceso analítico	257
evaluación de la calidad de un proceso vinculado a la salud	261
cálculo de la Métrica Sigma	258
selección del método	53-57, 210, 225
sensibilidad analítica	33, 49, 57, 208, 288
sensibilidad funcional	173, 175, 177, 182, 299, 310
sesgo	20-23, 32, 34, 43-44, 58-59, 66-67, 76, 79, 83-84, 88, 94-95, 97, 102, 125, 132-135, 140, 144-148, 150, 153-154, 160, 163, 168, 171, 176-177, 179, 190, 196-197, 199, 200-203, 212-213, 216, 225-226, 228, 232, 234, 241-244, 251, 253, 257-260, 269, 273-274, 276-280, 287-293, 295-297, 300, 303-304, 308-309
sesgo máximo permitido	22
sesgo permitido	20, 197, 278
sensibilidad	33, 43, 47-49, 57, 107, 139, 141, 148, 173, 175, 177, 182, 209-210, 267, 288, 299, 310
sensibilidad de las estadísticas a los distintos tipos de error	141
significación clínica	153-155
simulación de errores en el resultado de una prueba	141
sentido estadístico, sensibilidad y significancia	139, 141
sistemas de la calidad	39, 41, 281
soluciones de control	119
soluciones estándar	114, 119, 160, 165
solución interferente	159-161
suero	109, 119, 129
sustancias interferentes	48, 266
T	
tabla-F	98
tabla "t"	95, 122, 243, 246
tamaño de error crítico	22
Tonks	18
traduciendo las especificaciones de desempeño en una Métrica Sigma	253
transferencia	185, 187-189, 191-192, 211, 299, 301-302
transferencia de intervalos de referencia	187, 192
criterio divino	188
estimación a partir de 60 muestras	190
propósito	187
verificación a partir de 20 muestras	189
qué hacer	192
trazabilidad, valor trazable	34, 45, 58-59, 127, 263-264, 268-272, 276, 305
U	
uso previsto	57, 266-269, 272, 279
uso adecuado de las estadísticas	39, 221, 229
utilidad clínica	56, 81, 154, 176
V	
validación administrativa	9
validación clínica	9
validación del método	211, 256, 306
validación técnica	9
valores asignados	58, 81, 103, 108, 110-111, 114, 135, 164, 245, 247, 306
valor-F	290, 292-293
valor "t" calculado	153, 296
valor "t" crítico	153, 246, 249, 292
valor de verificación	239-240
valor target (diana)	34-35, 277, 279, 313
valores target (diana) asignados	34-35
valor verdadero	111-112, 246-247, 256, 270-271, 304
valores relativos	110
variabilidad biológica	6, 22-23, 278, 280

- variabilidad biológica entre individuos 23
- variabilidad biológica intra individuo 23, 278-280
- variación 6, 9, 19
- varianza 97, 122, 236-240, 246-247, 277
- varianza entre corridas 238-240, 247
- varianza intra corrida 238-240, 247
- varianza total 237-238
- veracidad 15, 34, 70, 127, 176, 231, 233-
 235, 241, 245-247, 263-265, 270, 273, 275, 277,
 279, 281, 305
- EP15 definición 241
- veracidad de la medición
 - definición ISO 270
- verificación de intervalos de referencia 185
- verificación de las especificaciones de desempeño
 41-42, 134, 239-240, 243, 245

Validación Básica de Método

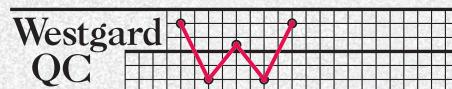
Revisado. Ampliado. *Esencial.*

Por más de 30 años, el Dr. James O. Westgard ha promovido estándares y buenas prácticas para la validación de métodos. En esta edición, la tercera edición de Validación Básica de Métodos, el Dr. Westgard y sus colegas brinda importantes actualizaciones sobre requisitos regulatorios, estándares de buenas prácticas emergentes y las ultimas guías de CLSI (Clinical and Laboratory Standards Institute):

- Actualización de las regulaciones de CLIA y requisitos de acreditación;
- Capítulo revisado sobre rango reportable que incluye una discusión sobre verificación de la calibración;
- Capítulo revisado sobre límites de detección que incluye los conceptos de Límite de Blanco, Límite de Detección y Límite de Cuantificación;
- “Carta de Decisión de Método” actualizada que incluye los criterios para un desempeño de dos, tres, cuatro, cinco, y seis sigma;
- Nuevo capítulo sobre la estimación de precisión y veracidad basado en la guía de las CLSI EP15-A2, incluyendo instrucciones para realizar los cálculos requeridos usando planillas electrónicas;
- Nuevo capítulo sobre estándares globales emergentes (ISO, International Standars Organization), el estándar ISO 15189, y los conceptos de veracidad e incertidumbre de medida.

Con una gran cantidad de material nuevo y actualizado, además de accesos online a herramientas para graficar y una base de datos sobre requisitos de la calidad, la tercera edición de Validación Básica de Métodos continua siendo una referencia práctica esencial para el laboratorio moderno.

Este libro ha sido traducido con la autorización de Westgard QC, Inc.



7614 Gray Fox Trail

Madison, WI 53717

Copyright 2008 Westgard QC, Inc.

<http://www.westgard.com>

Esta traducción fue posible
gracias a una donación de la
Fundación Wallace H. Coulter.