

Documento de opinión

Wytze P. Oosterhuis* and Elvar Theodorsson

Error total vs. incertidumbre de medición: ¿revolución o evolución?

DO 10.1515/cclm-2015-0997

Recibido el 8 de octubre de 2015; aceptado el 12 de octubre de 2015

Resumen: La primera conferencia estratégica de EFLM "Definición de objetivos de desempeño analítico, 15 años después de la Conferencia de Estocolmo" se realizó en otoño de 2014 en Milán. Mantuvo la jerarquía de objetivos de rendimiento de Estocolmo 1999, pero los reordenó, definió y estableció cinco grupos de trabajo para actuar en temas relacionados con los objetivos de desempeño analítico, incluido uno sobre la teoría del "error total". Jim Westgard escribió recientemente una visión general completa de los objetivos de desempeño y de la teoría del error total crítica de los resultados y las intenciones de la conferencia de Milán 2014. La teoría del "error total" originada por Jim Westgard y sus colaboradores tiene una influencia dominante en la teoría y la práctica de la química clínica, pero no se acepta en otros campos de la metrología. La teoría de la incertidumbre es generalmente aceptada, sin embargo, adolece de matemática compleja y no es concebible en la química clínica. Es necesario debatir los pros y los contras de la teoría del error total, dando paso a métodos que pueden incorporar todas las causas relevantes de incertidumbre al hacer diagnósticos médicos y monitorear los efectos del tratamiento. Este desarrollo debería proceder preferiblemente no como una revolución, sino como una evolución.

Palabras clave: error total permitido; objetivos analíticos de desempeño; error; error analítico total; incertidumbre.

Introducción

Todos los paradigmas y métodos científicos deben ser desafiados honestamente para desarrollarse y prevalecer. Este fue el objetivo de la primera conferencia estratégica de EFLM "Definición de objetivos de desempeño analítico, 15 años después de la Conferencia de Estocolmo" celebrada en otoño de 2014 en Milán. La conferencia de Estocolmo de 1999 constituyó un hito influyente en Química Clínica y la conferencia de Milán de 2014 honró a sus pioneros. Otros desarrollos recientes también contribuyeron a los incentivos para organizar la conferencia de Milán 2014, incluidos los estándares de acreditación ISO 17025 y 15189 que requieren que los laboratorios proporcionen rutinariamente la medición de la incertidumbre (MU) de los resultados de las mediciones y la armonización de la evaluación de procedimientos de evaluación de la calidad externa y desafíos recientes a la teoría del error total (ET) incluyendo el cálculo del error total permitido (ETa).

Los criterios de calidad de Estocolmo se mantuvieron en la conferencia de Milán 2014, pero se organizaron en paralelo en lugar de la jerarquía de Estocolmo 1999 [1]. Cinco Grupos de trabajo por un tiempo definido de EFLM (GTF) fueron iniciados por la Conferencia de Milán 2014 encargada de explorar y desarrollar los principales temas discutidos durante la Conferencia, entre ellos un grupo encargado de trabajar en ET de los cuales son miembros los autores actuales. Debido a que el trabajo en curso de los GTF no está terminado, las

opiniones expresadas en el presente documento representan solo las opiniones de sus autores y no el consenso del grupo.

El sitio web de Westgard (www.westgard.com) desde el principio criticó la conferencia de Milán 2014 comparándola con el edicto de Milán 313 AD. A pesar de su ominoso nombre, el edificio de Milán era en realidad un acuerdo entre el emperador romano occidental Constantino I y Licinio, que controlaba los Balcanes para tratar a los cristianos con benevolencia dentro del Imperio Romano. En contraste con la libertad plasmada en el consenso de Estocolmo 1999, el sitio web de Westgard expresó temor a la persecución y que aquellos que persisten en adherirse a los criterios originales de Estocolmo 1999 o al paradigma de ET se considerarán inmundos y se enfrentarán al ostracismo.

De hecho, la conferencia de Milán 2014 fomentó la apertura y la libertad de pensamiento en todos los temas de interés. Por lo tanto, los Westgards no tienen "nada que temer, sino temerse a sí mismos" que durante décadas han representado el desarrollo más influyente en el área del control de calidad en la química clínica.

Desde el principio, Sten Westgard ha sido y es un miembro bienvenido de EFLM y del grupo final en ET.

En un documento de opinión en el número actual [2] Jim Westgard llama la atención sobre varios temas que se discutieron durante la conferencia de Milán de 2014. Si bien en general es partidario de seguir trabajando en los temas, menciona enfáticamente las ventajas de los criterios originales de Estocolmo 1999 y del paradigma convencional de ET. El documento es loable en su cobertura integral de los principios de ET y ETa. Sin embargo, carece de perspectivas sobre las limitaciones del ET y ETa para acomodar todas las fuentes relevantes de incertidumbre encontradas en la química clínica. Nuestro propósito aquí es señalar áreas relacionadas con ET y ETa, donde encontramos que se necesita una nueva forma de pensar.

Los paradigmas de error e incertidumbre

El paradigma del error se inició a principios del siglo XIX i.a. por C.F. Gauss para dilucidar la verdadera información contenida en un resultado de medición. Los errores se dividieron en categorías de errores sistemáticos y aleatorios. Los errores sistemáticos pueden corregirse y en ocasiones eliminarse prácticamente mediante una calibración adecuada, mientras que los efectos de los errores aleatorios en la estimación de la media solo se pueden reducir, además de mejorar el método, mediante mediciones repetidas. Sin embargo, se hizo evidente que una gran cantidad de factores puede influir en el resultado de la medición y que el verdadero valor, la piedra angular del paradigma del error, nunca puede conocerse con exactitud. Por lo tanto, el paradigma de error que incluye a ET y ETa lucha cuando representa numerosos y complejos factores que influyen en los resultados de medición y comúnmente recurre a los modelos más simples de la realidad, por ejemplo; el concepto erróneo común entre los químicos clínicos de que la repetitividad y la reproducibilidad, como medidas de imprecisión a corto y largo plazo, de un resultado de medición es igual a su MU total.

El paradigma de MU, que se originó a principios del siglo XX [3], afirma que la incertidumbre es una propiedad de un resultado de medición que expresa la falta de conocimiento del verdadero valor del resultado e incorpora los factores que se sabe influyen en él. Para estimar la incertidumbre combinada de una medición, se deben conocer todas las fuentes de influencias aleatorias y sistemáticas y se debe estimar la contribución de cada componente principal a la UM mediante los métodos adecuados. MU ha sido generalmente aceptado en todos los campos de la metrología excepto en química clínica [4] a pesar de que

la Federación Internacional de Química Clínica y Medicina de Laboratorio (IFCC) es uno de los fundadores de métodos de incertidumbre en química [4, 5]

Error total y Error total permitido

El paradigma ET / ETa aspira a ser un modelo completo, pero adolece de una serie de deficiencias que se han vuelto particularmente evidentes últimamente.

En el paradigma convencional de TE, el sesgo y la imprecisión (multiplicados por un factor z) se añaden linealmente, lo que da como resultado un valor para el error total:

$TE = \text{bias} + zCVa$. Este valor representa una estimación del límite de un intervalo alrededor del valor "verdadero" (desconocido) donde los resultados analíticos medidos se pueden encontrar con una probabilidad definida (generalmente una probabilidad del 95%).

El mismo modelo se puede usar para derivar límites de error máximos (permitidos). Aquí encontramos una propiedad notable de la distribución gaussiana: cuando se define un límite con el número correspondiente de resultados de prueba fuera de este límite (por ejemplo, 5% en $z = 1,65$, unilateral) las combinaciones de sesgo e imprecisión (CVa) cumpliendo esta condición muestra una línea recta con una pendiente de -z. Por esta razón, el modelo se puede llamar lineal. En el extremo de la línea tenemos el sesgo = 0, con $ETa = 1.65CVa$, en el otro extremo tenemos $CVa = 0$ (un valor hipotético) y el sesgo = ETa.

El modelo de ET supone además que se conoce la diferencia ET entre el resultado del paciente y el valor verdadero, y si los resultados de los pacientes difieren menos de ETa del valor verdadero ($ET < ETa$), se satisfacen todas las necesidades de calidad clínicamente previstas.

Ninguna de estas suposiciones es, sin embargo, lo suficientemente comprensiva o adecuada para el propósito. Por lo general, no sabemos si ETa satisface la calidad prevista para el uso clínico de los resultados del paciente a pesar de que los "límites de aceptación" en muchas pruebas de aptitud / programas externos de garantía de calidad se expresan en forma de ETa. Estos criterios son, en muchos casos, técnicos y están establecidos para satisfacer los límites reglamentarios mínimos, como un cierto porcentaje de participantes que cumplen los criterios. En algunos casos, incluidos los CLIA de los EE.UU., los límites de rendimiento aceptables están consignados en la ley, sin documentación de su origen o su idoneidad para el uso clínico. Además, ciertos criterios pueden no ser aptos para el uso con fines de diagnóstico y monitoreo.

De acuerdo con el modelo ET / ETa lineal, la suma de un sesgo máximo y z veces una imprecisión máxima es igual a una constante, el ETa. El hecho de que el máximo sesgo e imprecisión se combinen con un número constante es muy conveniente en el modelo propagado por Westgard.

De acuerdo con este modelo, esta característica es válida sobre el rango total de imprecisión y sesgo. Esto permite el desarrollo de reglas de calidad para los procedimientos de control de calidad internos, independientemente de la relación de sesgo e imprecisión. Esto es de importancia crucial y fundamental en los métodos propugnados por J. Westgard y S. Westgard (www.westgard.com).

Sin embargo, este modelo solo es válido cuando la imprecisión y el sesgo son las únicas variables involucradas, o en otras palabras: cuando la distribución gaussiana supuesta se describe completamente por el sesgo de medición y la sola imprecisión. Este modelo, por ejemplo; no se ocupa de la variación biológica y otras causas evidentes adicionales de

variación. Este es un serio inconveniente de este modelo, ya que necesita incorporar posibilidades para lidiar con la variación en factores adicionales. La variación biológica no es relevante cuando se monitorea la variación en las muestras de control, pero ciertamente sí lo es cuando se trata de muestras de pacientes.

Los desarrollos en esta área son, por lo tanto, esenciales si los límites ETa se deben administrar para defender su uso actual generalizado.

El modelo convencional supone que ETa puede derivarse generalmente de la variación biológica. Los siguientes valores se usan comúnmente al calcular los límites ETa:

Sesgo: $<0.25(CV_i + CV_g)$; imprecisión $CV_a : <0.5CV_i$

Con CV_i = variación intraindividuo, CV_g = variación entreindividuos; CV_a = variación analítica.

Westgard defiende estas ecuaciones y constantes en su artículo. Es importante observar que estos valores representan valores máximos que resultan de teorías basadas en la influencia combinada del sesgo y la imprecisión en los valores de referencia [6]. Estos valores se derivaron en condiciones de exclusividad mutua: el término de imprecisión es válido solo bajo la condición de sesgo cero y, a la inversa, el término de sesgo solo es válido con cero imprecisión (una condición hipotética). Se ha demostrado que el uso de estas ecuaciones y constantes puede llevar a una sobreestimación seria de ETa [7].

Modelos mixtos y especificaciones de desempeño

Westgard se refiere a modelos alternativos como modelos mixtos, ya que tanto la variación analítica de última generación como la variación biológica se mezclan al calcular los límites ETa. Cuando se aplica un modelo basado en valores de referencia y valores de cambio de referencia (RCV, Reference Change Value), la variación analítica de "estado del arte" debe incluirse en el modelo. La razón es que el RCV está en parte determinado por esa variación analítica. Los procedimientos de control de calidad estarán dirigidos a mantener los métodos analíticos estables y los RCV válidos.

Deben distinguirse claramente dos fases en el ciclo de vida de una prueba. Durante la fase de validación, se establecen las diferentes medidas de rendimiento de una prueba y se toma la decisión de si cumple o no con su uso previsto. En una segunda fase donde la prueba está en uso rutinario, los procedimientos de control de calidad se establecen para mantener la calidad, incluyendo el mantenimiento de la validez de los valores de referencia, y en el caso de las aplicaciones de monitoreo, la validez de los RCV. Los modelos que estamos discutiendo aquí son apropiados para la segunda fase. El rendimiento de los métodos es conocido (y presumiblemente de última generación), y las medidas de calidad están destinadas a mantener este nivel. Incluir el estado de la técnica en el modelo no es un problema en este caso.

La relación de sesgo e imprecisión debe ser conocida para la estimación de ETa en los modelos mixtos [8] y los efectos de estos factores en el error total de la medición en teoría deben separarse. Sin embargo, esto solo puede hacerse en retrospectiva, y no en el caso de la interpretación de los últimos resultados del control de calidad interno. Esto puede resolverse suponiendo que la imprecisión sea constante, con un error adicional atribuido al sesgo [8].

Este supuesto, en parte, armoniza con el modelo de Westgard, donde las curvas de potencia suponen una imprecisión inicial estable.

El modelo de error total y la incertidumbre de medición

Los dos temas principales que favorecen el paradigma de MU son que la mayoría, si no todos los demás campos de la metrología, lo están utilizando y que alienta la estimación de los principales componentes de la incertidumbre y favorece las acciones para su minimización. GUM (Guía para la expresión de la incertidumbre) define la incertidumbre como "un parámetro, asociado con el resultado de una medición, que caracteriza la dispersión de los valores que podrían atribuirse razonablemente al mensurando". Este parámetro suele ser una desviación estándar o el ancho de un parámetro. intervalo de confianza. Esta incertidumbre se puede expandir por un factor (por ejemplo, 1.96): esto se interpreta como un intervalo en el cual el valor del resultado de una medición reside con una probabilidad definida. Con un factor de cobertura de 1.96, hay un 95% de probabilidad (bilateral) de que el resultado de la medición se encuentre dentro de este intervalo, suponiendo una distribución normal de los resultados [9].

Aunque se considera que el enfoque ascendente de la GUM no es bien aplicable para el laboratorio clínico, y se supone que la nueva directriz CLSI esperada no sigue GUM, varios conceptos de GUM son valiosos. GUM indica que todos los sesgos deben corregirse cuando sea posible. Esta es una de las contradicciones importantes dentro del modelo ET: ¿por qué mantener el error (sesgo), cuando se conoce el valor y se puede usar en el cálculo de ET? En la práctica de rutina, sin embargo, el sesgo a menudo no se conoce exactamente como la imprecisión y el sesgo no se pueden separar fácilmente. Por ejemplo, cuando tratamos de estimar el sesgo a partir de los resultados del control de calidad interno, el resultado dependerá del marco de tiempo: cuanto más amplio sea el marco de tiempo, más sesgo se promediará y se incluirá en el cálculo como imprecisión.

MU es útil por una serie de razones. Brinda información sobre la calidad de las mediciones; es útil para comparar la calidad metrológica de varios laboratorios clínicos (entre laboratorios clínicos acreditados, siempre que se calcule de la misma manera) y ayuda en la interpretación de los resultados de la medición, especialmente cuando se acercan a valores críticos (p. ej. definición de valores por patología, concentración de etanol en sangre para conductores, etc.). De hecho, cuando se compara un resultado con un límite de decisión (por ejemplo, 7,0 mmol / l en el diagnóstico de diabetes), podemos proporcionar información clara al médico solo si el límite no se incluye dentro de la incertidumbre en torno al resultado. Entonces, no hay duda de que el concepto es valioso. Fuera del laboratorio, podemos encontrar el modelo MU, cuando las medidas de velocidad de la policía de tráfico se comparan con los límites de velocidad solo después de una corrección basada en la incertidumbre de la medición.

¿Son las diferencias entre ET y MU reconciliables?

Las diferencias y similitudes de MU y ET han sido revisadas [9-11]. Estos modelos tienen sus diferencias, pero no se debe enfatizar demasiado. ET v.g. define una región alrededor del valor de referencia ("verdadero") donde los resultados analíticos medidos se pueden encontrar con una probabilidad definida mientras que MU define una región alrededor del resultado analítico medido donde se puede encontrar el valor "verdadero" con una

probabilidad definida. La similitud entre ambos modelos es que ambos expresan la confiabilidad del resultado de la prueba, sin embargo, desde una perspectiva diferente.

Cuando el modelo MU se hace lo más simple posible eliminando el sesgo y solo tratando con pocas causas independientes de variación, los dos modelos se vuelven matemáticamente idénticos, esto es, controlan la calidad interna de un único método donde se encuentran muchas fuentes de variación que son excluidas al medir muestras de pacientes. En la práctica, las causas de variación son mucho más complejas, incluyendo, variación dentro del método, entre los métodos, dentro y entre los sistemas de medición, dentro y entre los laboratorios, dentro y entre los lotes de reactivos y las calibraciones, dentro y entre los individuos y la variación preanalítica.

En su artículo [2], Westgard intenta de manera apreciable reconciliar el error analítico total y la MU. Considera "error analítico total [ETa] como una medida de precisión y MU como medida de trazabilidad" y cree que "ETa es esencial para gestionar la calidad en un laboratorio médico y MU y la veracidad son esenciales para lograr la comparabilidad de los resultados de los laboratorios ". Consideramos que esta solución es demasiado simplista. Los métodos de ET / ETa deben tener en cuenta las complejas realidades atendidas por los métodos de MU y las últimas deben desarrollarse en la realidad práctica que los métodos ET / ETa han tenido en cuenta en la química clínica durante décadas. Los modelos MU se han usado principalmente para estimar la confiabilidad de los resultados de medición. La aplicación para establecer objetivos no está bien desarrollada. Sin embargo, en el control interno de la calidad, los modelos ET / MU se vuelven más o menos iguales, y la fijación de objetivos desarrollada en el modelo ET es aplicable para ambos. Sin embargo, debe tenerse en cuenta que los modelos desarrollados para estimar el sesgo máximo permisible y la imprecisión basados en valores por RCV no se basan teórica ni prácticamente en ninguno de los dos modelos y pueden aplicarse a ambos.

La incertidumbre de medición se incluye inherentemente en el resultado de la medición y, como tal, se expresa en valores de referencia (cambio) y estimaciones de sensibilidad y especificidad y otras medidas de precisión diagnóstica. Los procedimientos de control de calidad destinados a controlar la variación analítica basados en el control de calidad interno no capturarán todas las fuentes de variación. Solo los procedimientos que incluyen todas las fuentes de variación nos permitirán comprender la MU total real al medir muestras de pacientes [12].

Conclusiones

Los conceptos de ET y ETa originados por Jim Westgard y sus colaboradores han servido bien a la química clínica durante décadas y todavía representan una influencia dominante en la teoría y la práctica de la química clínica. El paradigma de la incertidumbre es ampliamente aceptado en otros campos de la metrología, pero ha sufrido de matemáticas complejas y la impracticabilidad concebida. ET no está reconocido ni en el Vocabulario Internacional de Metrología (4) ni en el GUM (5) que favorece a MU. Sin embargo, los paradigmas ET y MU pueden aparecer como dos caras de la misma moneda. Sus ecuaciones que describen la variación, pueden convertirse en lo mismo cuando se elimina el sesgo y pocas causas independientes de variación están en juego. Sin embargo, su filosofía fundamental es diferente y tiene diferentes consecuencias en las prioridades prácticas de la química clínica.

Ni la separación propuesta por Westgard ni la sustitución de ETa con MU resolverán las tensiones entre ETa y MU dentro de la química clínica. En nuestra opinión, los modelos ET / ETa implementados en química clínica se pueden mejorar sustancialmente mediante la implementación de estimaciones de componentes de incertidumbre y métodos para cálculos de incertidumbre ya aplicados en otras disciplinas dentro de la metrología. Los desarrollos iniciales en esta dirección se están implementando actualmente en la química clínica [10, 11, 13].

Los pros y los contras de los conceptos ET y ETa deben debatirse, dando lugar a métodos que puedan incorporar todas las causas relevantes de incertidumbre al hacer diagnósticos médicos y monitorear los efectos del tratamiento. El cálculo de la incertidumbre de nuestros resultados representa una nueva e importante oportunidad para mejorar la calidad de los servicios de química clínica para el diagnóstico y la monitorización de los efectos del tratamiento de los pacientes.

Para ser productivo al máximo, este desarrollo debería proceder preferiblemente no como una revolución sino como una evolución del concepto de "error total". Sin embargo, tal evolución depende de la libertad de lealtad a todos los aspectos de los paradigmas que ya han cumplido su propósito. Esta es una lección adecuada para aprender del edificio de Milán 313 AD.

Contribuciones de los autores: Todos los autores han aceptado la responsabilidad del contenido completo de este manuscrito enviado y la presentación aprobada. Apoyo financiero: ninguno declarado.

Empleo o liderazgo: Ninguno declarado.

Honorario: Ninguno declarado.

Conflicto de intereses: las organizaciones de financiación no desempeñaron ningún papel en el diseño del estudio; en la recopilación, análisis e interpretación de datos; en la redacción del informe; o en la decisión de enviar el informe para su publicación.

Referencias (Se decidió dejar en formato original para que se mas fácil si posterior consulta)

1. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: Consensus Statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. *Clin Chem Lab Med* 2015;53:833–5.

2. Westgard JO. Useful measures and models for analytical quality management in medical laboratories. *Clin Chem Lab Med* 2015. Sep 30 pii: /j/cclmahead-of-print/cclm-2015-0710/cclm-2015-0710xml doi: 10.1515/cclm-2015-0710. [Epub ahead of print].

3. Peat FD. From certainty to uncertainty: the story of science and ideas in the twentieth century. Washington, DC: Joseph Henry Press; 2002:xiv, 230.

4. JCGM. Evaluation of measurement data – guide to the expression of uncertainty in measurement. JCGM 100:2008, GUM 1995 with minor corrections: Joint Committee for Guides in Metrology; 2008. Available at: http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf. Accessed 24 October, 2015.

5. JCGM. Evaluation of measurement data – Supplement 1 to the “Guide to the expression of uncertainty

in measurement” – Propagation of distributions using a Monte Carlo method. Joint Committee for Guides in Metrology: 2008. Available at: [http:// www.bipm.org/en/publications/guides/](http://www.bipm.org/en/publications/guides/). Accessed 24 October, 2015.

6. Gowans EM, Hyltoft Peteresen P, Blaabjerg O, Hörder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988;48:757–64.

7. Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. *Clin Chem* 2011;57:1334–6. 8. Oosterhuis WP, Sandberg S. Proposal for the modification of the conventional model for establishing performance specifica-

tions. *Clin Chem Lab Med* 2015;53:925–37. 9. Rozet E, Marini RD, Ziemons E, Dewé W, Rudaz S, Boulton AJ,

et al. Total error and uncertainty: friends or foes? *Trend Anal*

Chem 2011;30:797–806. 10. Farrance I, Frenkel R. Uncertainty of measurement: a review of the rules for calculating uncertainty components through func-

tional relationships. *Clin Biochem Rev* 2012;33:49–75. 11. Farrance I, Frenkel R. Uncertainty in measurement: a review of monte carlo simulation using microsoft excel for the calculation of uncertainties through functional relationships, including uncertainties in empirically derived constants. *Clin Biochem Rev* 2014;35:37–61.

12. O’Donnell GE, Hibbert DB. Treatment of bias in estimating measurement uncertainty. *Analyst* 2005;130:721–9.

13. WhiteGH, FarranceI, GroupAUoMW. Uncertainty of measurement in quantitative medical testing: a laboratory implementation guide. *Clin Biochem Rev* 2004;25:S1–24.