

Sr. Director:

He leído con interés el trabajo que bajo el título "Validación de un termómetro auricular instantáneo por infrarrojos" y cuyos autores son J. Korta, A. Alberdi y J.I. Empanza aparece publicado en esta revista⁽¹⁾.

En dicho trabajo se desea comprobar el grado de acuerdo o concordancia (*agreement*) entre un termómetro instantáneo por infrarrojos con el instrumento de referencia, el termómetro de mercurio. En este tipo de estudios lo que generalmente interesa es conocer si ambos reproducen o no, el mismo resultado. Es decir, si son intercambiables.

Los estudios pueden ser de: a) Consistencia (*reliability*): si ninguno de los dos procedimientos es considerado ideal, la concordancia mide el grado de consistencia entre ambos procedimientos; b) Conformidad (*accuracy*): si un procedimiento es comparado con otro considerado ideal, la concordancia mide el grado de conformidad de aquél con este. En general, la mayoría de los índices de concordancia se usan para describir tanto la consistencia como la conformidad.

Otros conceptos a tener en cuenta son: a) Precisión (fiabilidad o reproductibilidad): un aparato es preciso si en múltiples determinaciones del mismo parámetro en un mismo sujeto en condiciones idénticas proporciona valores muy similares. La precisión cuantifica el error aleatorio derivado de las determinaciones repetidas; b) Exactitud (validez): el valor esperado de medidas repetidas coincide con el verdadero valor. La inexactitud, se define también como error sistemático o sesgo (bias). Es la capacidad de un instrumento para medir aquello para lo que ha sido diseñado. La exactitud de un aparato indica el grado en que la medida se aproxima al valor real (patrón oro). No tiene sentido evaluar la exactitud de un aparato poco preciso.

En el trabajo de Korta y cols. la metodología utilizada es la comparación de la t de Student apareada, el análisis de la varianza, la correlación lineal de Pearson y el análisis de Bland-Altman o método de las medias de las diferencias.

En el trabajo se observa como los autores ponen el acento en la comparación de la t de Student apareada y ofrecen unos valores de significación con diferencias estadísticamente significativas entre la medida de la temperatura auricular con respecto a la temperatura rectal y a la oral (la lectura de la tabla es confusa). Es bien conocido que las diferencias estadísticamente significativas no tienen por qué ser clínicamente relevantes y

Consideraciones sobre la validación de un termómetro auricular instantáneo por infrarrojos

así en dicha tabla se observa cómo entre la temperatura auricular y la rectal existe una diferencia de 0,2°C y entre la temperatura auricular y la oral de 0,3°C que en clínica no se puede considerar importante.

Se muestran los valores de la correlación lineal de Pearson con cifras de 0,8 entre la temperatura rectal y auricular y de 0,9 para la temperatura axilar y auricular. El coeficiente de correlación mide la intensidad de la asociación entre dos variables, no su concordancia. Es obvio que dos métodos diseñados para medir una misma variable deben tener, cuando menos, cierto grado de asociación. Una correlación perfecta no garantiza la concordancia. Si los métodos concuerdan en sus medidas, sus valores serán intercambiables y, por tanto, un método puede sustituir al otro⁽²⁾.

Para poder valorar el grado de concordancia se debe utilizar el coeficiente de correlación intraclass, que permite valorar la información de la variación de los sujetos, la variación sistemática entre los métodos y, por último, la variación aleatoria entre mediciones⁽³⁻⁶⁾. En este trabajo no se menciona este método que es considerado fundamental para este tipo de investigaciones.

El análisis de Bland-Altman completa el estudio anterior por cuanto de forma sencilla y gráfica se puede evaluar la concordancia entre dos métodos de medida prefijando unos límites de concordancia con sentido eminentemente clínico^(2,6-7). Pero debe ser realizado de forma correcta. De esta manera, se obtienen las medias de las diferencias entre cada método así como los intervalos de concordancia con los límites de confianza correspondientes de forma que a intervalo más estrecho, mayor grado de concordancia.

En el trabajo de Korta se muestra un nivel de significación para la prueba de Bland-Altman ($p < 0,001$) del que desconocemos su significado, puesto que en este análisis no existe ningún tipo de contraste ni nivel de significación. Además, se muestran dos gráficas confusas con líneas de regresión lineal que no corresponden a esta metodología.

Sería deseable que la revista Anales Españoles de Pediatría realizara revisiones en el aspecto metodológico y que también se analizaran desde este punto de vista los trabajos recibidos.

Bibliografía

- 1 Korta Murúa J, Alberdi Alberdi A, Empanza Knörr JI. Validación de un termómetro auricular instantáneo por infrarrojos. *An Esp Pediatr* 1998; **48**: 195-197.
- 2 Candela Toha AM. Valoración de aparatos y métodos de medida:

- concordancia sí, correlación no. *Med Clin (Barc)* 1992; **99**: 314.
- 3 Fleiss JL. The design and analysis of clinical experiments. New York: John Wiley and Sons Inc; 1986.
 - 4 Bland JM, Altman DG. Measurement error and correlation coefficient. *BMJ* 1996; **313**: 41-42.
 - 5 Prieto L, Lamarca R, Casado A. La evolución de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase. *Med Clin (Barc)* 1998; **110**: 142-145.
 - 6 Gómez de la Cámara A, Cruz Martos E. Análisis de la fiabilidad de tres dosímetros portátiles de glucemia. Comparación de diferentes métodos para el estudio de la fiabilidad de las observaciones clínicas. *Med Clin (Barc)* 1997; **108**: 410-413.
 - 7 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **1**: 307-

J. Korta, A. Alberdi, J.I. Emparanza

An Esp Pediatr 1999;50:104-105.

Réplica

Sr. Director:

Agradecemos los comentarios del Dr. Andrés a nuestra carta. Coincidimos con él en algunos de sus comentarios sobre estudios de concordancia de mediciones clínicas, que por otra parte, son tratados en diversos textos de bioestadística básica, donde se incluye también el caso de concordancia entre variables categóricas.

En el caso que nos ocupa, de medición y variables cuantitativas, los estudios de concordancia son muy frecuentes en la clínica y pocas veces se realizan con una metodología correcta. De modo práctico, deben incluir tres fases:

1. Comparación pareada de las mediciones. Esto nos da una idea, por una parte, de la magnitud de las diferencias y de su relevancia clínica y, por otra, de la significación estadística.

2. Análisis de correlación. Si mediante este análisis concluimos que no existe asociación, el estudio de concordancia se puede dar por finalizado. La existencia de asociación no garantiza la concordancia, pero es un requisito imprescindible para que sea posible.

3. Análisis de la concordancia en sí mismo. Existen diversas formas de realizarlo, como explicamos a continuación.

El primero de ellos es el cálculo del coeficiente de correlación intraclase, preferido por el Dr. Andrés. Ciertamente, la obtención de una cifra que puede variar entre 0 y 1 (los valores negativos son poco probables en la práctica), expresando falta total de acuerdo (concordancia) o acuerdo perfecto, parece atractiva. Dado que existen varios coeficientes de correlación intraclase (CCI), tanto paramétricos como no paramétricos, el primer problema es cuál elegir.

El primero, publicado por Fisher en 1925, asume igualdad

de medias y de varianzas, y es equivalente a un ANOVA de una vía, en el que se descompone la varianza en intrasujeto e intersujeto. Posteriormente se empleó el ANOVA de dos vías, ya sea considerando a los observadores como una muestra aleatoria o fija. En estos dos métodos la varianza intrasujeto se descompone en variabilidad interobservadores, interacción entre observadores y sujetos (mediciones) y un término error. Por lo tanto, estos dos modelos permiten diferenciar entre sesgo y error aleatorio. Tras los trabajos de Fleiss, fue descrito otro CCI utilizando la suma de cuadrados de ANOVA en lugar de los cuadrados medios. Este CCI depende directamente del tamaño muestral, por lo que ha sido criticado. El último CCI paramétrico descrito se compone del coeficiente «r» de Pearson y un factor corrector del sesgo, que mide la distancia entre las mediciones y la línea de identidad. Además de estos cinco CCI paramétricos citados, se han descrito otros cuatro CCI no paramétricos.

El uso de los CCI paramétricos puede, en general, ser criticado, porque pueden ocurrir violaciones de las condiciones de aplicación o asunciones: distribución normal multivariante e igualdad de varianzas. Se ha demostrado también que cualquier CCI paramétrico depende del rango de la escala de medición. En general, cuanto más amplio es el rango mejor CCI se obtiene.

Del mismo modo, el uso de los CCI no paramétricos tiene sus desventajas. En primer lugar, ningún paquete estadístico permite realizar su cálculo de modo sencillo. Por otra parte, estos CCI son poco sensibles a la presencia de «outliers». Por último, presuponen la intercambiabilidad de las mediciones de los observadores, asunción que frecuentemente no es cierta en la clínica.

Dado que los CCI no son una medida satisfactoria, debido a su dependencia de la variación intersujeto y a que combinan en una sola medida los dos tipos de errores que pueden afec-

Consultorio Médico. Donostia, San Sebastián.

tar al acuerdo entre métodos de medición (sesgos y errores aleatorios), Altman⁽¹⁾ propuso un nuevo método (alternativo, no complementario). Mediante la media de las diferencias se cuantifica el sesgo, mediante la desviación estándar de las diferencias se cuantifica el error aleatorio, y mediante los límites de acuerdo se predice la máxima diferencia esperable entre los dos métodos.

No creemos, por tanto, que el cálculo del CCI sea un método fundamental como afirma el Dr. Andrés. Pese a que en el pasado lo hemos utilizado^(2,3), creemos que su cálculo da una falsa seguridad al clínico, y que existen muchas limitaciones para que sea considerado como fundamental. En su lugar abogamos por el uso del método de Altman, debido a sus ventajas: es sencillo de realizar y de entender. Dada su sencillez, no entendemos la objeción de que debe ser realizado de forma correcta: en primer lugar se presentan las medias de cada par de valores frente a su diferencia. Por costumbre, se colocan las medias en el eje de las abscisas y las diferencias en el de ordenadas. Se dibuja una línea horizontal en el valor de la media de las diferencias y otras dos en los valores correspondientes a esta media más y menos 1,96 desviaciones estándar. El siguiente paso, de significado desconocido para el Dr. Andrés, consiste en un estudio de regresión lineal entre los valores medios y las diferencias, tomando esta última como variable dependiente. Debe calcularse y mostrarse el nivel de significación. Niveles de significación menores al 5% indican una tendencia a variar (aumentar o disminuir) la diferencia entre las mediciones a lo largo del rango de medición. Las gráficas obtenidas, lejos de ser confusas, son fácilmente enten-

dibles por el clínico con una somera explicación.

Parece evidente que, lejos de recomendaciones simplistas, los estudios de concordancia exigen conocer bien cuál es el propósito del estudio (evidenciar sesgos, conocer influencia de «outliers», etc.). Si la elección es el cálculo de un CCI, debemos ser cuidadosos en su elección, conociendo exactamente el tipo de muestreo al que hemos accedido. Frente al cálculo del CCI preferimos el análisis gráfico de Altman, que permite descubrir «outliers» y examinar tendencias (sesgos) mediante un análisis de regresión lineal. Su inconveniente⁽⁴⁾ es que no proporciona una cifra numérica única y que, por lo tanto, la evaluación del acuerdo debe recaer finalmente en el juicio clínico. Para nosotros este último inconveniente, lejos de serlo, permite al clínico huir del criterio estadístico, frecuentemente rígido, e incorporar su propia maestría y experiencia clínica en la decisión. Una ventaja añadida importante es que puede ser fácilmente entendido por personas sin formación estadística.

Bibliografía

- 1 Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983;**32**:307-317.
- 2 Areses R, Empanaza JI, Arriola M, Urbieto MA. Estudio haurtxo. Valores de referencia de la calcemia y de la calciuria en nuestra población infantil normal. *Calcio en la infancia. Nefrología* 1994;**14**:584-590.
- 3 González E, Mintegui J, Garmendia A, Callén MT, Empanaza JI. Los «monitores de bolsillo» de la función respiratoria: ¿son útiles en niños? *An Esp Pediatr* 1994;**41**:388-390.
- 4 Lee J. Evaluating agreement between two methods for measuring

J. Díez Domingo, A. Ballester Sanz

An Esp Pediatr 1999;**50**:105-106.

Sr. Director:

Una práctica muy extendida entre los pediatras, tanto de Atención Primaria como de los servicios de urgencias hospitalarios, es la administración secuencial de dos antitérmicos, de forma que cada dos a cuatro horas el niño recibe un antitérmico distinto.

Centro de Salud de Nazaret. Valencia.

Correspondencia: Javier Díez Domingo. C.S. Nazaret. C/ Parque, 7. 46024 Valencia.

Utilización secuencial de antitérmicos. ¿Basado en la evidencia o en la costumbre?

Esta pauta es llamativa, ya que en pediatría es excepcional la utilización de dos fármacos como primera elección para un proceso, y cuando se hace, por ejemplo con los antibióticos, se utilizan sustancias con mecanismos de acción diferentes y sinérgicos. Los tres fármacos más utilizados en nuestro medio para el tratamiento de la fiebre, el ácido acetil salicílico (AAS), el paracetamol y el ibuprofeno, tienen, sin embargo, un mecanismo de acción común, que es la inhibición de la ciclooxigenasa.

Para localizar las pruebas científicas que apoyen la utiliza-