

Programa de evaluación de la calidad entre laboratorios. XXII. La confiabilidad de la media de consenso.

Elvia M. Cabañas-Cortés,¹ Rosa María Sánchez-Manzano,² Sergio I. Alva-Estrada¹

Abstract

External evaluation of quality control programs checks accuracy of laboratories analyzing the same sample. Establishment of target value is the most critical stage in this evaluation. This value can be set up by reference laboratories, consensus of participants or by preparing controls with high purity components.

We studied the reliability of consensus in 16 evaluations, comparing the values obtained using identical serums. We calculated the error of variance rate punctuation (PIV) between mean values, it was an average value of 26 and it is equivalent to the points difference that is 6.5 % and 25 % of 400 and 100 points, which represent the maximum and limit values respectively. Error is higher in components with high PIV and variation coefficient average values and vice versa. In spite of error magnitude the consensus mean can be used and it is much better when the interlaboratories variability is low.

Keywords: *Target value. Consensus mean. Quality Control. External Quality Control.*

Resumen

Los programas externos evalúan la exactitud de los laboratorios que analizaron la misma muestra. El paso crítico de la evaluación es el establecimiento del valor blanco o meta, que puede ser definido por laboratorios de referencia, por consenso de los participantes o por preparación de controles a partir de componentes puros.

Se estudió la confiabilidad del consenso en 16 evaluaciones, comparando los valores obtenidos con sueros idénticos. Se calculó el error de la puntuación del índice de varianza (PIV) entre las medias, y el promedio fue 26 que equivale a la diferencia en puntos, o sea 6.5% y 25% de las cifras de 400 y 100 puntos, que son los valores máximo y límite del PIV, respectivamente. El error es mayor en componentes cuyos valores del PIV y coeficiente de

variación interlaboratorio son altos y viceversa. A pesar de la magnitud del error la media de consenso se puede utilizar, y es más confiable cuanto menor variabilidad interlaboratorios haya.

Palabras Clave: *Valor blanco. Media de consenso. Evaluación Externa de la Calidad. Control de Calidad.*

Introducción

En los programas de evaluación externa de la calidad (EEC) se evalúa la exactitud de resultados reportados por los laboratorios después de que analizaron muestras problema de un mismo lote. El paso crítico para esto es la asignación del valor considerado como blanco o meta, contra el cual se compararán dichos resultados. Para ello existen diferentes opciones: **a)** que sea establecido o asignado por laboratorios de referencia que analizaron los materiales problema; **b)** que se considere como valor esperado al promedio aritmético de los resultados obtenidos por los participantes, al que se denomina media de consenso; y **c)** que se conozca la concentración de cada componente porque el material analizado se prepara por adición de componentes puros. Sin embargo cada uno de estos métodos tiene sus desventajas.

En el primero, el costo es muy elevado ya que los laboratorios de referencia deben analizar la muestra un número de veces significativo, utilizar métodos definitivos, de referencia, y los que los laboratorios emplean comúnmente para establecer en cada uno las concentraciones de los analitos. Otro inconveniente es que difícilmente se prueban todos los métodos y marcas de reactivos disponibles.

La media de consenso se establece reuniendo los resultados de los laboratorios que emplean cada método, excluyendo los datos extremos y calculando el promedio de los datos restantes. El problema es que se necesita un número considerable de resultados, para que los datos extremos no afecten.⁽¹⁾ A pesar de esta limitante se han informado buenos resultados en un programa con 60 laboratorios⁽²⁾ y en otro con 35.⁽³⁾

La desventaja de utilizar controles preparados por adición es que no tienen muchos componentes, por problemas de solubilidad y, en consecuencia no están sujetos a todos los efectos de matriz que se presentan en los

¹ Profesores del Departamento de Bioquímica, ENCB-IPN; becarios de COFAA y del Programa de Estímulo al Desempeño Docente.

² Profesora del Departamento de Parasitología, ENCB-IPN.
Correspondencia: Dr. en C. Sergio I. Alva Estrada. Departamento de Bioquímica. Escuela Nacional de Ciencias Biológicas. Prof. de Carpio y Plan de Ayala. Colonia Santo Tomás. 11340 México D.F.

materiales de origen natural, como los porcinos, equinos y humanos.

Los efectos de matriz son un problema común a los tres sistemas descritos, a ellos se debe que, para un mismo analito, se acepte que hay concentraciones diferentes, dependiendo del método o marca utilizada.

En estudios comparativos de los valores asignados por laboratorios de referencia y las medias de consenso, se han encontrado discrepancias, algunos autores prefieren los primeros,^(4,5) sin embargo las medias de consenso se aplican en programas de EEC de diferentes países del mundo,⁽⁶⁾ como Portugal,⁽⁷⁾ Italia,⁽⁸⁾ España,^(9,10) Jordania,⁽¹¹⁾ Reino Unido⁽¹²⁾ e Israel⁽³⁾ y se aplican con éxito desde 1969, así mismo en el United Kingdom External Quality Assessment Scheme (UKEQAS), que ha sido adoptado por diferentes programas de otros países.^(10,13) También se ha informado su aplicación, con buenos resultados, en hematología.⁽¹⁴⁾

Para Whitehead⁽¹³⁾ el uso del método de la media es correcto por las siguientes razones: **a)** la experiencia de años es que las medias de consenso y los valores establecidos por laboratorios de referencia son semejantes, **b)** la media y la moda para las determinaciones son virtualmente las mismas, indicando iguales efectos positivos y negativos sobre el valor de la media de consenso, y **c)** para una misma sustancia, los diferentes métodos analíticos dan también medias de consenso idénticas.

A pesar de lo antes señalado, algunos autores⁽¹⁵⁾ consideran que en los esquemas de evaluación externa de la calidad (EEC), no se debe asumir a la media de consenso como el valor blanco, sin evidencias de su validez, obtenidas por medio de otros estudios.

Desde que se inició el Programa de Evaluación de la Calidad Entre Laboratorios (PECEL), en octubre de 1990,

Tabla 1. Estudio de la confiabilidad de la media de consenso en la evaluación externa de la calidad analítica. Para cada suero y componente se señala el coeficiente de variación seleccionado o error permitido (CVS), el promedio de las medias de consenso (MC) y su coeficiente de variación, el promedio de las puntuaciones del índice de varianza (PIV) y su error (PIVerror), obtenidos en cuatro evaluaciones en que se uso el mismo lote de suero.

COMPONENTE	CVS	SUERO QUALITROL DE NIVEL 1				SUERO QUALITROL DE NIVEL 2			
		MC*	CV**	PIV**	PIVerror	MC*	CV**	PIV**	PIVerror
GLUCOSA	7.7	85.0	1.76	60	23	245.0	1.03	75	13
UREA	5.7	43.0	1.16	121	20	100.0	1.76	132	31
CREATININA	8.9	1.4	1.36	69	11	3.3	1.22	82	12
ACIDO URICO	7.7	5.5	1.22	79	14	8.8	0.83	82	10
PROTEINA TOT	3.9	6.5	0.92	93	20	6.5	1.04	57	22
ALBUMINA	7.5	4.1	1.21	59	14	4.0	1.12	58	13
BILIRR. TOTAL	10.0	1.3	1.61	83	12	2.7	2.78	80	23
BILIRR. DIRECT	10.0	0.3	6.87	119	26	0.6	5.85	149	32
COLESTEROL	7.6	125.0	1.40	87	18	125.0	2.29	94	30
TRIGLICERIDOS	7.6	84.0	2.21	121	29	90.0	3.00	117	39
AST	10.0	58.0	1.55	82	15	180.0	1.74	72	17
ALT	10.0	36.0	1.56	107	15	125.0	1.69	74	17
ALP	10.0	150.0	3.60	122	36	380.0	3.52	117	35
LDH	10.0	200.0	2.65	90	26	300.0	1.99	106	20
AML	10.0	120.0	3.32	136	33	300.0	3.57	141	36
CKNAC	10.0	200.0	3.59	96	36	430.0	2.30	103	23
Na	4.0	142.0	0.48	61	12	165.0	0.55	65	14
K	4.0	4.8	0.48	69	10	6.6	0.99	64	21
Cl	4.0	100.0	1.50	80	37	118.0	1.21	80	30
Ca	4.0	8.5	1.95	133	43	11.7	2.04	117	46
P	4.0	4.4	2.06	140	40	6.8	1.68	130	35
Fe	8.0	102.0	2.16	94	27	190.0	1.03	76	13
PROMEDIO				96	23			94	24

* Corresponde al promedio de las medias de consenso del método más utilizado.

** Corresponde al promedio de todos los métodos utilizados.

se utiliza con éxito la media de consenso, y en este trabajo se estudia su confiabilidad, comparando los valores obtenidos en varias ocasiones, en que se empleó un mismo lote de sueros problema.^(16,17,18)

Material y Métodos

Para la evaluación mensual de la exactitud de los laboratorios,⁽¹³⁾ éstos analizan sueros de un mismo lote, informan los resultados y métodos utilizados a los organizadores. Posteriormente se reúnen los resultados o valores observados (V.obs.) de acuerdo al método o sistema analítico utilizado, se establece la media de consenso y por comparación con ésta, se determina la inexactitud de cada participante a través del cálculo de la puntuación del índice de varianza (PIV).

La media de consenso, valor esperado, blanco o meta se establece en tres etapas: 1) del conjunto de resultados

recibidos de los laboratorios participantes y agrupados de acuerdo al método/instrumento utilizado, se eliminan aquellos que se salen del intervalo de 20 a 200% de la media aritmética obtenida (primer media). 2) con los datos restantes, se establece una segunda media, a la que se suma y resta su propia desviación estándar y 3) con los datos incluidos entre esos límites se establece una tercer media, que es considerada como la media de consenso o valor esperado (V.esp.).

Para calcular el PIV, primero se determina el % de error mediante la siguiente fórmula:

$$\%Error = (V.obs. - V.esp.) \times 100/V.esp.$$

Posteriormente se obtiene el PIV dividiendo el % de error resultante entre el coeficiente de variación seleccionado (CVS) o % de error aceptable y multiplicando por 100. Los CVSs se señalan en las tablas 1 y 2.

Tabla 2. Estudio de la confiabilidad de la media de consenso en la evaluación externa de la calidad analítica. Para cada suero y componente se señala el coeficiente de variación seleccionado o error permitido (CVS), el promedio de las medias de consenso (MC) y su coeficiente de variación, el promedio de las puntuaciones del índice de varianza (PIV) y su error (PIVerror), obtenidos en cuatro evaluaciones en que se uso el mismo lote de suero.

COMPONENTE	SUERO BIORAD DE NIVEL 1					SUERO BIORAD DE NIVEL 2			
	CVS	MC*	CV**	PIV**	PIVerror	MC*	CV**	PIV**	PIVerror
GLUCOSA	7.7	88.0	1.43	64	18	281.0	1.14	65	15
UREA	5.7	38.0	1.36	135	23	100.0	2.23	122	39
CREATININA	8.9	1.7	2.72	138	23	5.8	1.87	100	19
ACIDO URICO	7.7	4.8	1.39	88	16	9.2	2.04	82	25
PROTEINA TOT	3.9	5.9	0.91	98	19	4.3	1.09	98	22
ALBUMINA	7.5	3.6	1.48	74	17	2.7	1.74	62	19
BILIRR. TOTAL	10.0	1.0	3.09	106	21	4.8	2.07	81	19
BILIRR. DIRECT	10.0	0.4	7.60	112	34	1.6	3.74	109	28
COLESTEROL	7.6	120.0	1.94	86	25	90.0	2.12	85	28
TRIGLICERIDOS	7.6	200.0	2.33	102	31	100.0	2.91	107	38
AST	10.0	28.0	2.13	99	21	170.0	4.65	78	46
ALT	10.0	30.0	3.03	109	30	90.0	1.97	89	20
ALP	10.0	45.0	3.20	157	32	300.0	3.55	139	35
LDH	10.0	150.0	1.30	136	13	350.0	2.43	127	24
AML	10.0	100.0	5.10	187	51	300.0	5.60	161	56
CKNAC	10.0	***				***			
Na	4.0	146.0	0.35	53	9	130.0	1.10	61	27
K	4.0	4.1	0.52	67	10	6.4	1.76	66	37
Cl	4.0	100.0	1.98	91	49	90.0	1.69	97	42
Ca	4.0	8.0	1.36	139	29	11.4	1.41	100	32
P	4.0	2.7	3.34	181	57	7.5	1.85	133	40
Fe	8.0	241.0	4.69	76	58	***			
PROMEDIO				109	28			98	30

* Corresponde al promedio de las medias de consenso del método más utilizado.

** Corresponde al promedio de todos los métodos utilizados.

*** No se reunió el número de datos necesario.

Los valores del PIV pueden variar desde 0 a 400. La calidad se considera satisfactoria cuando está entre 0 (ideal) y 100 (límite), regular de 101 a 150, mala de 151 a 200 y pésima con valores mayores de 200.

Para evaluar la confiabilidad de la media de consenso, se utilizó el mismo lote de sueros control en cuatro evaluaciones, experiencia que se repitió con cuatro sueros distintos, correspondientes a dos marcas (Qualitrol y Biorad) y dos niveles de concentración de cada una (N1 y N2). En las evaluaciones participaron en promedio 400 laboratorios, que con sus métodos habituales, determinaron los 22 analitos que incluye el programa. En las 16 evaluaciones involucradas, los participantes desconocían el lote de suero utilizado.

Se reunieron las 4 medias de consenso de cada componente, de cada uno de los cuatro lotes de suero utilizado y se calcularon, de manera habitual, los

promedios, desviaciones estándar y coeficientes de variación (CV); para el cálculo del error del PIV, el coeficiente de variación resultante se multiplicó por 100 y se dividió entre el CVS. El error del PIV corresponde al nivel de incertidumbre que se tendría, si se utilizaran las diferentes medias de consenso de un mismo suero. Cabe señalar que el resultado numérico tiene las mismas dimensiones que el PIV, y equivale a la diferencia en puntos, que se obtendría con esas medias de consenso. Con los resultados de una evaluación y para cada método se calculó el CV interlaboratorios.

Resultados

En las tablas 1 y 2 se muestran los resultados obtenidos para los sueros Qualitrol y Biorad en los dos niveles de concentración utilizados, respectivamente. En ambos casos y para cada uno de los 22 componentes que se evalúan en el PECCEL, se incluyen los valores de CVS

Tabla 3. Comparación de los coeficientes de variación interlaboratorios obtenidos en el PECCEL y en programas de evaluación externa de la calidad de otros países.^(9,10)

COMPONENTE	PECCEL*	CAP	UKEQAS	SFBC	DGKCH	SEQC
GLUCOSA	3.1-12.2					
UREA	4.1-29.4					
CREATININA	4.6-26.7	5.3	5.9	6.0	10.0	9.7
ACIDO URICO	3.0-17.0					
PROTEINA TOT	3.9-12.9					
ALBUMINA	5.0-15.2		5.9	6.5		7.6
BILIRR. TOTAL	3.9-24.9					7.6-24.3
BILIRR. DIRECT	9.2-56.1					
COLESTEROL	4.3-14.4					
TRIGLICERIDOS	5.4-18.9	5.9		7.9	15.0	9.5
AST	4.8-33.8		12.5	8.4	14.0	11.6
ALT	3.8-29.3					
ALP	6.1-54.0		24.2	14.4	13.0	19.7
LDH	3.9-28.8			34.8	13.5	28.4
AML	8.3-40.7					
CKNAC	8.3-35.1					24.7-54.1*
Na	3.3-16.8					
K	3.8-16.6					
Cl	2.4-80.0					
Ca	4.2-31.8					
P	4.2-20.1					
Fe	3.2-30.9					

* Se presenta el CV más bajo y alto, de los diferentes métodos o sistemas analíticos.

PECCEL	Programa de Evaluación de la Calidad Entre Laboratorios
CAP	College of American Pathologists
UKEQAS	United Kingdom External Quality Assessment Scheme
SFBC	Société Française de Biologie Clinique
DGKCH	Deutsch Gesellschaft für Klinische Chemie
SEQC	Sociedad Española de Química Clínica

utilizados para el cálculo del PIV, los promedios de cuatro evaluaciones de: las medias de consenso (se presenta sólo la del método más utilizado); los coeficientes de variación; del PIV y del error estimado del PIV (PIVerror).

En la tabla 3 se presenta para cada componente, el intervalo de los coeficientes de variación interlaboratorios, que se obtuvieron con los diferentes métodos utilizados por los participantes en el PECEL. Se incluyen también los datos equivalentes disponibles, de programas de evaluación externa de otros países.

Discusión

En la tabla 1 se puede apreciar que los promedios del error del PIV, en los sueros Qualitrol de niveles de concentración 1 y 2, son de 23 y 24 respectivamente, cifras prácticamente idénticas entre sí y semejantes a los valores de 28 y 30, correspondientes a los sueros Biorad de los mismos niveles (tabla 2). Estos datos señalan así, que en promedio se podría llegar a tener un error de 26 puntos, en la evaluación de un laboratorio, dependiendo del valor de la media de consenso que se obtuviera.

El error de 26 puntos de PIV es pequeño, si se considera que representa sólo un 6.5% de los 400 puntos, que es su valor máximo, aunque puede parecer significativo si se considera que representa el 26% de la cifra límite de 100 puntos.

En las tablas 1 y 2 se aprecia también, que en general, el error del PIV es mayor en aquellos componentes cuyo promedio del PIV es alto. En el fósforo por ejemplo, con el suero Biorad de niveles 1 y 2 se obtuvieron promedios de PIV de 181 y 133 puntos, con un error de 57 y 40, respectivamente.

Por otro lado, en la tabla 3 se puede apreciar que el intervalo de los coeficientes de variación interlaboratorios es muy amplio, por ejemplo, el de fósforo es de 4.2% a 20.1%. Es pertinente señalar que el primero corresponde a métodos o equipos automatizados y el segundo a métodos o sistemas manuales, de manera que el error del PIV, que depende directamente de la variabilidad en la media de consenso, puede estar relacionado con el CV interlaboratorios.

A pesar de la magnitud del error, los resultados sugieren que la media de consenso se puede seguir utilizando, sobre todo porque es más confiable cuanto menor es la variabilidad interlaboratorios.

Los intervalos de los CVs interlaboratorios obtenidos en el PECEL (tabla 3), en general son muy amplios y abarcan los obtenidos en programas de evaluación de otros países, de los que lamentablemente no se cuenta con los intervalos correspondientes, situación que permitiría comparar la calidad alcanzada en diferentes países. Puede señalarse sin embargo, que los valores más pequeños de

CV interlaboratorios son menores que los de otros programas, lo que sugiere que en México tenemos laboratorios que trabajan con una calidad óptima, situación que deberá generalizarse al mejorar la calidad de los métodos o sistemas, cuyos promedios del PIV y CVs interlaboratorios sean altos, o bien al aplicar métodos o sistemas con los que ya se ha demostrado exactitud.

Referencias

1. Bullock DG, Smith NJ y Whitehead TP. External Quality Assessment of Assays of Lead in Blood, *Clin Chem* 1986; 32:1884-9.
2. Georges RJ. Validity of the consensus mean as the target value for a small External Quality Assessment Scheme, *Ann Clin Biochem* 1985;22(Pt 3):283-90.
3. Kahana L y Yechieli H. What can be achieved in an external quality-assessment scheme with a small number of participants: four years of experience with thyroid-related test in Israel, *Clin Chem* 1991;37(8):1432-6.
4. Hansert E y Stamm D. Determination of assigned values in control specimens for internal accuracy control and for interlaboratory surveys. Evaluation of 200 different lots with identical experimental design: experiences and conclusions, *J Clin Chem Clin Biochem* 1980;18:461-90.
5. Uldall A, Blaabjerg O, Eikving S, Elg P, Gerhardt W, Holmberg H, Horder M, Icen A, Juva K, Jorgensen PJ: A programme for assigning target values for external quality assessment schemes in countries with no authorized reference laboratories. Annex. Experiences with deviating results on Ektachem 700 XR, *Scand J Clin Lab Invest Suppl* 1993;212: 31-7.
6. Akinyanju PA. Quality Control in Developing Countries, *Clin Chem News* 1981;131-3.
7. Martins MC, Rodrigues MO, Fonseca A. Quality assurance programmes in Portugal in the fields of clinical chemistry and haematology. *Ann Ist Super Sanita* 1995;31:37-41.
8. Albertini A, Signorini C. The quality assurance system in clinical chemistry, *Ann Ist Super Sanita* 1995;31:3-8.
9. Ramón F, Alsina MJ, Cortés M, Navarro JM, Ricós C y Salas A. Evaluación del X Programa de Control de Calidad Interlaboratorios de la SEQC, *Quím Clín* 1990; 9:137-221.
10. Ramón F, Alsina MJ, Alvarez V, Cortés M, Hernandez A, Jiménez CV, Michinela J, Navarro JM, Perich C, Ricós C, Salas A y Simón M. Evaluación del XI Programa de Control de Calidad Interlaboratorios de la SEQC, *Quím Clín* 1991;10: 133-97.
11. Bilto YY. External quality assessment of Jordanian clinical chemistry laboratories, *Ann Clin Biochem* 1992;29(Pt 3):324-30.
12. Seth J, Hanning I, Bacon RR y Hunter WM. Quality of performance of assays for serum growth hormone in the United Kingdom (UK): evidence from the external quality assessment scheme, 1980-1987, *Clin Chim Acta* 1988; 174:171-83.
13. Whitehead TP. Quality control in clinical chemistry. "Advances in Quality Control", John Wiley and Sons, New York, 175-205.
14. Vives-Corrons JL, Gutiérrez G, Jou JM, Reverter JC, Martínez-Brotos F, Domingo A, Iriarte JA. Characteristics and performance of the external quality assessment scheme (EQAS) for haematology in Spain. Ten years of experience, *Ann Ist Super Sanita* 1995;31:95-101.
15. Seth J, Sturgeon CM, Ellis AR, al Sadie R. Validation of target values in external quality assessment schemes for peptide hormones and tumour markers, *Scand J Clin Lab Invest Suppl* 1993;212:28-30.
16. Curiel-López P, Fuentes-Mancilla LM, Cabañas-Cortés EM, Lara-Uc M, Alva-Estrada SI, Valles de Bourges V, González-S MA y Romero-Camarena A. Programa de evaluación de la calidad entre laboratorios. VII. Resumen de resultados de dos años, *Lab-acta* 1993;5:37-42.
17. Alva-Estrada SI, Fuentes-Mancilla LM, Valles de Bourges V, Romero-Camarena A, Lara-Uc M y Olivia-Arana BA. Programa de evaluación de la calidad entre laboratorios. XI. Resultados de cuatro años, *Lab-acta* 1994;6:131-7.
18. Alva-Estrada SI, Fuentes-Mancilla LM, Cabañas-Cortés EM y Sánchez-Mancilla RM: Programa de evaluación de la calidad entre laboratorios. La Calidad en química clínica dentro de límites XVII, *Lab-acta* 1998;10:31-6.